DOCUMENT RESUME

ED 284 703 RC 016 325

AUTHOR Hackett, Eugene de G.; McManus, James V.
TITLE An Intra Population Analysis of Migrant Student Performance Based on the Migrant Student Record Transfer System Data Base.
SPONS AGENCY Arkansas State Dept. of Education, Little Rock.
PUB DATE 25 Aug 86
NOTE 304p.; Some tables contain small print.
PUB TYPE Reports - Research/Technical (143) -- Statistical Data (110)

EDRS PRICE MF01/PC13 Plus Postage.
DESCRIPTORS Academic Achievement; Age Differences; Age Grade Placement; Comparative Analysis; Databases; Elementary Secondary Education; Instructional Program Divisions; *Mathematics Achievement; *Migrant Education; *Online Systems; *Program Effectiveness; *Reading Achievement; *Skill Analysis; Test Norms
IDENTIFIERS Impact; *Migrant Student Record Transfer System

ABSTRACT

Educational services data contained in the Migrant Student Record Transfer System (MSRTS) database were used to examine some effects of the migrant education program on student performance, and how student performance varied with the amount and duration of exposure to general and particular program components. Problems with using the MSRTS database for research were the absence of control groups and non-continuity of student data and the program. Reading, Language and Arithmetic sections of the Comprehensive Tests of Basic Skills and Reading and Arithmetic sections of the California Achievement Test scores were used to divide data into two groups: below grade and at/above grade. Data were analyzed using simple polynomial, multiple regressions, analysis of variance, and the General Least Squares Method. Independent variables were entry age, migrant status, grade level, test type, and reading, mathematics and arithmetic skills. Fifteen experimental sets were organized to measure various aspects of the program. Some findings were that migrant status, test types, and grade level produced significant differences in student performance; and entry age and grade level were highly correlated. Appendices include a discussion of the variables, population, tests, procedures, and results for each set; information on the population and samples; and an instructions manual containing statistical tables. (JMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED284703

AN INTRA POPULATION ANALYSIS OF MIGRANT STUDENT PERFORMANCE BASED ON THE MIGRANT STUDENT RECORD TRANSFER SYSTEM DATA BASE

TEXT

by
Eugene deG. Hackett
James V. McManus

August 25, 1986

RC016325

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Winfred Miller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

ACKNOWLEDGEMENTS

The authors wish to recognize the support provided by the entire MSRTS technical staff. Special thanks are due the operations and systems groups who managed to accommodate our needs despite the usual overwhelming work loads.

Dr. Alan Johnson, Chairman, Department of Mathematics and Statistics, University of Arkansas at Little Rock provided the steady hand of expertise in matters statistical. Dr. Johnson's immediate insight into the heart of the numbers, his skill in communicating and his patience and involvement were an inspiration.

Appreciation goes to Mr. Winford "Joe" Miller for understanding that thoroughness and schedules are natural enemies.

A NOTE ON CAUSALITY

The authors wish to emphasize a precept that is sometimes given cursory mention and then ignored. As statistical techniques, regression and correlations are close kin. Both address the co-occurrences of two or more variables. Neither technique "knows nor cares" about the feasibility of any variable being a "causal agent" for other variables. The statistical analyses simply provide humans the opportunity to display wisdom and commit folly in making assertions about causality, and such assertions usually contain a mixture of both of these ingredients. It is the proportions that matter! The authors recognize that "effects" attributed to various aspects of the migrant education program may, in fact, be the result of many other factors. It is less obstructive to communications to make this acknowledgement initially than it is to sprinkle endless qualifications throughout the discussions of results.

ORGANIZATION OF DOCUMENTATION

The documentation for this study consists of two separate documents. One document contains the text, and the second document contains tables and plots that could not be integrated into the text without requiring that the user skip back and forth between text and figures. The authors hope that the present organization will be convenient when referring to the various figures discussed in the text.

The text is organized into major parts on the basis of the subjects under investigation:

general program effects
math skill system effects
reading skill system effects
appendix.

BEST COPY AVAILABLE

Recognizing that some readers may be interested in one of these subject areas but not others, each part has been written as a stand alone document as long as Part I (Study Design) and the Appendix are recognized as companions.

TABLE OF CONTENTS
PART I
STUDY DESIGN

		PAGE
A.	Objective	I-1
B.	Problem Statement	I-1
	1. Problems of Using the MSRTS Data Base for Research	
	2. Problems of Measurements in Educational Research	
C.	Approach	I-2
	1. Group Comparisons	
	2. Index of Student Performance	
	3. Sample and Sample Sizes	
	4. Analysis Techniques	
	5. General Experimental Design	
D.	General Procedures	I-4
E.	Definitions of Key Terms Used in the Study	I-5
F.	Pseudo Hypotheses	I-6

PART II
GENERAL EFFECTS OF THE MIGRANT EDUCATION PROGRAM
ON STUDENTS
ARITHMETIC, READING AND LANGUAGE
STANDARDIZED TEST SCORES

A.	EXPERIMENTAL SET A	II-1
	Main effects of entry age, status, test type and grade level.	
B.	EXPERIMENTAL SET B	II-3
	Regression of scores at or above grade on entry age within status, test type, grade level.	
C.	EXPERIMENTAL SET C	II-7
	Relationship between entry age and grade level.	

D. EXPERIMENTAL SET D II-10

Main effects of grade level, status and test type on test performance.

E. EXPERIMENTAL SET E II-12

Regression of scores at or above grade on grade level within status, test type, entry age.

F. EXPERIMENTAL SET F II-15

A comparison, at each grade level, of migrant students having minimal exposure to the migrant program with migrant students who entered the migrant program at age six.

G. EXPERIMENTAL SET G II-18

Joint effect of entry age and grade level on student performance.

PART III
EXAMINATION OF THE MATH SKILLS SYSTEM
AND
ITS RELATIONSHIP TO STUDENT PERFORMANCE

H. EXPERIMENTAL SET H III-1

Descriptive statistics of math skills data in the SKILLS data set.

I. EXPERIMENTAL SET I III-7

Descriptive statistics of math topics represented by the math skills data in the SKILLS data set.

J. EXPERIMENTAL SET J III-10

Descriptive statistics of math skills data and math test data in IDSHTDSK. For clarity the data set IDSHTDSK will be called "the Skills/Test data set."

K. EXPERIMENTAL SET K III-26

Migrant student performance on standardized math tests as a function of varying numbers of math skills in MSRTS records. The data set used for this analysis was IDSHSK which for clarity will be called "the Skills/Test data set."

PART IV
EXAMINATION OF THE READING SKILLS SYSTEM
AND
ITS RELATIONSHIP TO STUDENT PERFORMANCE

L. EXPERIMENTAL SET L IV-1

Descriptive statistics of reading skills data in the SKILLS data set.

M. EXPERIMENTAL SET M IV-5

Descriptive statistics of reading levels represented by the reading skills data in the SKILLS data set.

N. EXPERIMENTAL SET N IV-7

Descriptive statistics of reading skills data and reading test data in IDSHTDSK. For clarity, the data set IDSHTDSK will be called "the Skills/Test data set."

O. EXPERIMENTAL SET O IV-21

Migrant student performance on standardized reading tests as a function of varying numbers of reading skills in MSRTS records. The data set used for this analysis was IDSHSK which for clarity will be called "the Skills/Test data set."

APPENDIX

A. Populations and Samples

B. Populations Selected for Use in Studying the General Effects of the Migrant Education Program on Standardized Scores in Math and Reading

C. Populations Selected for Use in Studying the Effects of the Skill System on Standardized Scores in Math and Reading

D. SAS Analysis Module References

E. Glossary of Terms

TABLE OF CONTENTS

PART I
STUDY DESIGN

		PAGE
A.	Objective	I-1
B.	Problem Statement	I-1
	1. Problems of U.using the MSRTS Data Base for Research	
	2. Problems of Measurements in Educational Research	
C.	Approach	I-2
	1. Group Comparisons	
	2. Index of Student Performance	
	3. Sample and Sample Sizes	
	4. Analysis Techniques	
	5. General Experimental Design	
D.	General Procedures	I-4
E.	Definitions of Key Terms Used in the Study	I-5
F.	Pseudo Hypotheses	I-6

PART I-1

A. Objective

The MSRTS data base contains educational and health services data on thousands of migrant students. This study uses the educational services data to examine some of the effects of the migrant education program.

B. Problem Statement

1. Problems of Using the MSRTS Data Base for Research

Although the MSRTS contains hundreds of thousands of data items on migrant students, its purpose is not that of research. The data that MSRTS collects, stores and transmits is used to support educational programs for the migrant students. Suitability of MSRTS data for research or program evaluation purposes is strictly fortuitous. Since the MSRTS data base was designed for operational support of field level educational programs rather than for research, it presents many formidable obstacles to using its data for research or program evaluation.

The three most frequently encountered problems in using the MSRTS for research are:

*the absence of control groups.
The only subjects on the MSRTS data base are migrant students. Attempts to assess effects of the migrant education programs by comparing the performance of migrant to regular students cannot be done by using the MSRTS data base only. One may attempt to do so by comparing the "scores" of migrant students against nationally "normed" scores. Such comparisons, however, cannot in any meaningful way represent an experimental vs. control group comparison unless equivalency of groups has been established empirically. The establishment and maintenance of such equivalencies would constitute a sizeable and expensive task.

*non-continuity of student data.
Even though the MSRTS data base contains thousands of scores from nationally used tests (CTBS, CAT, etc.), attempts to measure changes in migrant student performance (however measured) over time meet with serious discontinuities of data. The mobility of the migrant student produces irregular schedules of testing and assessment of student achievement that frustrate identification of "growth," "gain," etc.

Even the final outcome of migrant students' educational experiences cannot be assessed from the MSRTS data base. The reliability of

notification by schools of dropouts, graduations or other terminations of educational endeavors is unknown and cannot be determined from the MSRTS data base.

*non-continuity of program.
Though national in scope, the migrant educational program is not monolithic. Migrant students may encounter programmatic differences between schools that render longitudinal student performance data meaningless.

The latter two problems are part of an underlying problem in attempting to use the MSRTS as a source of data for "evaluating" educational programs. This underlying problem is simply that the MSRTS data base is an incomplete representation of the educational experience of the migrant student. Of course, any single source of data (surveys, one-dimensional studies, testing programs, etc.) suffers this same research debility.

2. Problems of Measurement in Educational Research

For each conclusion reached concerning the effectiveness of an educational program, non-trivial objections may be raised. Controversy abounds concerning the core of educational studies. The proper measurement of "gains," the selection of "unbiased and relevant" tests, identifying meaningful intervals over which to measure retention of results are but a few of the measurement issues inherent in assessment of educational programs. The problems encountered when assessing regular education programs are further compounded by the variety of nationalities, ethnic backgrounds, languages and other socio-economic factors represented by the migrant population.

Much of the design of the analyses making up this study is in direct response to these measurement problems and to the availability of data in the MSRTS data base.

C. Approach

1. Group Comparisons

All comparisons were made between and within groups of migrant students to counter the problems of:

a. comparing dissimilar groups, i.e., migrant vs. non-migrant,
b. biases in the dependent variable test scores,
c. the absence of data on non-migrant students on the MSRTS data base.

This approach to group comparisons permits migrant students to serve as their own control groups.

CTBS: Reading, Language and Arithmetic
CAT: Reading and Arithmetic.

CTBS and CAT scores were dichotomized into two classes; below grade and at/above grade. Percent of scores below grade vs. percent at/above grade were then computed as group scores for the various comparisons.

3. Sample and Sample Sizes

In all cases, all data meeting selection criteria for membership in the groups being compared were used. No attempts were made to draw random samples within group selection criteria. There were three reasons for this decision:

*The migrant student population is very heterogeneous. Identifying bias in proposed simple (i.e., non-stratified) random sampling plans would in itself entail substantial study.

*The alternative to simple random sampling would be stratified sampling. The difficulty with this approach is one of defining strata that would assure representative samples. If, for example, only the prima facie classifications of ethnicity, mobility, stream and linguistic ability were selected for selecting a stratified sample, the study would probably require the entire data base just to populate each stratum.

*The purpose of the study is to learn whatever possible about effects of various aspects of the migrant education program on student performance. All information which might assist in attaining this purpose is of value. With contemporary techniques of computer supported analysis, one must justify not using available information--especially in view of so much uncertainty regarding the underlying characteristics of the population.

Use of the entire data base must not tempt one to believe that the study dealt with the "population" of migrant student data rather than a "sample." The data base that existed when the data were formed into analyzable "data sets" represented a snapshot in time. That particular data base will never exist again. The data used in the study has and will be preserved for examination by anyone so interested and authorized. However, the inferences made in the course of this study have value only to the extent that the "snapshot" used in the study represents such snapshots of the migrant data base occurring within a reasonable time zone. For a more orderly and detailed discussion of the populations and samples used for this study, please read "Populations and Samples," Section C or the Appendix.

4. Analysis Techniques

The SAS data analysis package was used for all analyses. This package made it possible to use much more of the data than would otherwise have been possible because of the ease of weighting values for unequal numbers of cases. The principal analyses used were simple, polynomial and multiple regression and analysis of variance. The particular models used were of the General Least Squares Method (GLM).

5. Experimental Design

The general experimental design used throughout was to compare the effect on student performance (being below or at/above grade) of varying degrees (amounts, durations) of exposure to general and particular components of the migrant education program. Other major independent variables investigated were grade level, migrant status, test type, age of entry into the program and presence of skills data in student records. Where the availability of repeated observations permitted, interactions among independent variables were investigated.

D. General Procedures

From the MSRTS data base, several data sets (files) containing principal accumulations of data that would be needed for analysis were created. From these data sets, data was selected to represent various populations of interest to the study. The following is a brief description of these data sets and their uses:

****
****
** The reader is encouraged to read Sections A, B and C **
** of the Appendix to understand fully the formation and **
** content of data groupings used in the analyses whose **
** results are the substance of this report. **
****
****

PART II that follows reports a number of analyses of the general effects of the migrant education program on student performance. For these analyses, a collection of migrant student records were selected from the data base using the criterion that

> at least one CTBS or CAT test scores from the years 1983-1985 must be present.

The collection of records so selected is named the "General Effects File." As records were selected for this file, an additional data item, "entry age" (the age of the student when first enrolled in any migrant program), was calculated and added to each record selected. A total of 351,011 records were included in the General Effects data set.

PART III that follows reports a number of analyses of the effects of skills on student performance. To perform these analyses, the following three files were constructed of records drawn from the data base:

i. A Test Data Set consisting of test data records for those students who had CAT or CTBS tests reported as having been administered during the years 1983-1985. This data set contains 144,148 records representing 351,011 tests.

ii. The Skills Data Set consisting of the skills data records for all students who had skills data reported from the inception of the skills system through 1985. This file contains 656,067 records representing 8,830,809 skills.

iii. The Skills and Test File consisting of the skills data records and the test data records for each student having a record in the Test Data Set described above. This file is composed of 144,148 records representing 1,090,834 skills and 351,011 tests.

There are many populations of data investigated in the course of this study. The formation and content of each of these populations is described in detail in the Appendix. References to these population descriptions are made throughout the discussion of each analysis. The reader is urged to review this material in the Appendix to the point of a firm understanding of the population under study.

E. Definitions of Key Terms

There are several terms used in this study whose exact meaning is not established by a working knowledge of the migrant education program. These terms and the meaning intended herein are given in the following:

1. Student Performance

Unless otherwise stated, this term means the percent of a group whose test scores are at/above grade level.

2. Entry Age

The age of a migrant student when first enrolled in a migrant education program as reported to the MSRTS.

3. Exposure to the Migrant Education Program

The length of time elapsed between a student's first enrollment in a school as a migrant student and the experimental cutoff time being used for comparison purposes. The quantity, exposure, is a unit of time

only and implies neither intensity nor continuity of exposure nor quality of the program to which the student is exposed. Exposure may be an inferred quantity such as when groups of different entry ages are compared.

4. Pseudo Hypotheses

Each of the major independent variables investigated in this study is discussed as a "Pseudo Hypotheses." Actual hypotheses used for decision rules, though technically correct, reveal little of the investigative intent at hand. In the following, each Pseudo Hypothesis (abbreviated PH) represents a liberalized statement of a hypothesis regarding the effects of the independent variable under discussion. Although a Pseudo Hypothesis may be directional, its technical counterpart need not be.

5. Experimental Sets

The statistical analyses were performed in sets. Each set investigated the effects of the independent variable of principal interest (and its relationship to the other independent variables) on a particular population. Each separate analysis is referred to as an experiment. Thus, each experimental set is a series of experiments whose primary purpose is to increase understanding of the principal variable and whose secondary purpose is to identify other contributing variables.

F. Psuedo Hypotheses

(See section, E. Definition of Key Terms, preceding for a definition of pseudo hypotheses)

Variable=Exposure (as reflected by entry age).
PH1=Groups with more exposure to the migrant education program will perform better than groups with less exposure.

The first general hypothesis investigated by the study concerned the effect of entry age on student performance. The rationale for the hypothesis would be as follows. If the migrant education program has a positive effect, then student performance will increase with increased exposure to the program. Entry age is used to anchor the beginning of exposure. Some event such as the occurrence of a test score at a given grade level establishes the upper bound of an exposure duration.

Variable=Entry age (per se)
PH2=Groups with earlier exposure to the migrant education program will perform better than groups with equal duration but later exposure.

Many studies have shown that the impact of educational programs varies with student age, the earlier the age the greater the impact. This known relationship makes necessary the evaluation of a second hypothesis (H2) regarding entry age.

Entry age, program exposure and grade level are linked in an almost mechanical way. In general, for migrant students, the following relationships among these three variables hold in the vast majority of cases.

Entry age	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Grade	(preschool)	1	1	2	3	4	5	6	7	8	9	10	11	12	12	12	12				
Exposure (1)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

(1) Exposure shown is maximum. Actual exposure may be obtained from the expression (current age - entry age).

The above chart specifies the inter-dependencies among the three variables, entry age, grade and exposure. For example, comparing either different entry age groups or different grade level groups may inadvertently involve comparing groups with different amounts of exposure.

Variable=Grade level
PH3=Groups tested at lower grade levels will perform better than groups with equal exposure tested at higher grade levels.

Critical to the current study is the recognition that for groups, the two variables, entry age and grade, have a fixed relationship. If, for example, two different entry age groups are compared using their first year scores then, by necessity, the test results of two different grade levels are being compared. As will be seen, this relationship between entry age and grade level is always significant and requires the introduction of a third hypothesis. Prior experiences with programs for educationally disadvantaged show that the higher the grade level the greater the performance discrepancy between the disadvantaged and regular students. Accordingly, pseudo hypothesis PH3 is introduced.

Variable=Status (migrant status 1, 2 and 3)
PH4=Student performance will not vary by status.

Migrant status is also a variable that may be determined by other independent variables. For example, the classification of a status 3 normally means that a migrant student has been in the program for at least a year. Therefore, it is not meaningful to speak of a group of status 3 migrant students with zero (0) program exposure. In addition

to these kinds of "subtle" inter-dependencies, there are two more factors which make the effects of status difficult to assess or control.

These factors are:

*The reporting of any data other than enrollment data declines sharply upon assignment of status 3.

*Changing a students classification from a status 1 or a status 2 to a status 3 is a post hoc administrative act which has no real time parallel in the world of the student. The effects of "settling out" may range from profound to none and may be delayed for years or may begin to show immediately.

In view of the foregoing considerations, status is included as an independent variable as a precaution only. Statements regarding the effect of status on student performance are more suspect than those regarding any of the other variables being investigated.

Variable=Test type
PH4=Student performance will vary by test type.

Difference between group scores may vary in three ways:

a. by test publisher (i.e., CAT vs. CTBS),
b. by subject area within each publisher (e.g., arithmetic vs. reading vs. language for CTBS scores) and
c. by subject area across publishers (i.e., arithmetic scores, regardless of publisher, may differ from reading scores, regardless of publisher).

Although this variable was not of major interest to the present study, it was investigated to control its effects, if any, on the major variables of interest.

Variable=Skills
PH5=Student performance will vary by the number of skills reported.

The goal of the MSRTS skills system is to support enrolling schools in maintaining continuity of instruction by their use of information concerning the individual migrant student's most recent area of instruction in math, reading and language. Unfortunately this goal is often frustrated by the mechanics of communication--both in receiving skills data from prior schools and in "triggering" the transmission of skills data from the MSRTS. This problem is a complex interaction of economic, privacy of information, technological, organizational and other factors with which the migrant education program at large has struggled for a long time. It is a problem whose effective solution requires, among other things, direct access of the MSRTS data base by

each classroom teacher. Until the program can afford such an extensive communications capability, it attempts to cope with the separation between the consumer of information (the classroom teacher) and the data (the MSRTS data base) with a variety of techniques, procedures, organization and devices.

As things now stand, two timed events must both transpire before a classroom teacher receives skills data from a prior school in time for it to be of use in maintaining continuity of instruction. The first event is for the sending school to transmit the skills data to the MSRTS data base prior to the student's enrollment in the next school. The second event is for the receiving school to "trigger" the flow of student data to the school immediately upon a student's arrival at the school (i.e., via a data request or an enrollment).

From perusal of transaction time lapse data, it is reasonable to assert that the probability of either event is low. If the two events are (statistically) independent then the joint occurrence of both events for an individual is equal to the product of the probability of occurrence of both events individually (which is to say equal to or less than the event having the least probability).

All the above considerations conspire to produce a cycle of non-stimulus and non-response that works to the detriment of the skills system. The classroom teacher does not receive skills data in time to incorporate it into the instructional planning for the individual student and therefore has never come to rely on the skills data for that purpose. The teacher cannot, then, be expected to be an ardent supporter of the skills system and cannot be expected to extol the virtues of the skills to the various levels of program administrators. The program administrators, hearing little unsolicited direct support of the skills system by teachers, are reluctant to devote additional program resources to enhancing the acceptance, use and development of the skills system. This is the perfect illustration of a viscous cycle. There is no blame, yet everyone loses--especially the student.

Despite all the factors working against the skills system, and in keeping with the spirit of this study which is outcome rather than process oriented, the effect of skills on student performance will be examined.

Notice that PH5 makes no prediction of the direction of the relationship between students having skills in their MSRTS records and their performance on tests. Nor does PH5 imply that a constant relationship exists--that is, the direction and magnitude of the relationship may vary by grade level or by frequency of skills.

TABLE OF CONTENTS

PART II

EXPERIMENTAL SETS A-G

GENERAL EFFECTS OF THE MIGRANT EDUCATION PROGRAM
ON STUDENTS
ARITHMETIC, READING AND LANGUAGE
STANDARDIZED TEST SCORES

		PAGE
A.	EXPERIMENTAL SET A	II-1
	Main effects of entry age, status, test type and grade level.	
B.	EXPERIMENTAL SET B	II-3
	Regression of scores at or above grade on entry age within status, test type, grade level.	
C.	EXPERIMENTAL SET C	II-7
	Relationship between entry age and grade level.	
D.	EXPERIMENTAL SET D	II-10
	Main effects of grade level, status and test type on test performance.	
E.	EXPERIMENTAL SET E	II-12
	Regression of scores at or above grade on grade level within status, test type, entry age.	
F.	EXPERIMENTAL SET F	II-15
	A comparison, at each grade level, of migrant students having minimal exposure to the migrant program with migrant students who entered the migrant program at age six.	
G.	EXPERIMENTAL SET G	II-18
	Joint effect of entry age and grade level on student performance.	

****
* All tables and plots referenced herein will be found in the *
* accompanying ILLUSTRATIONS MANUAL. *
****

Note 1: When inspecting the analyses of variance tables presented herein, always refer to the TYPE III values (as opposed to TYPE 1).

Note 2: In essence, the data plots represent populations. Vertical differences between data points may often appear not to be significant when in fact they are.

Note 3: Simple regressions are normally run for A on B and B on A. For this reason, the independent/dependent variables involved in the regression analyses are simply specified as members of the set (A, B).

Note 4: All testing is done at alpha = .05.

A. EXPERIMENTAL SET A

Main effects of entry age, status, test type and grade level.

1. Variables:

   Independent-Entry age, status, test type, grade level.
   Dependent-Percent of students scoring at or above grade level.

2. Population:

   General effects data set. Each record contains a test type. For each test type, each record included a related: test outcome, status, entry age and grade level. Number of observations = 2,859. Number of cases = 351,011.

3. Test:

   Analysis of variance using GLM main effects model.

4. Hyp:Ho:

   Main effect means of independent variables are equal.

   Hl:

   Main effect means of independent variables are not equal.

5. Procedure:

   Since no repeated measures were available, a main effects analysis of variance model with no interactions was run. This model tests for differences among the levels within each independent variable (e.g., status 1 vs. status 2 vs. status 3, etc.), but it does not test independent variables against each other.

6. Results:

   All main effect differences were significant. See Table Al.

7. Discussion of Results:

   The main thrust of this portion of the study is to examine the effect of the migrant program on student performance. Entry age is used as an indirect measure of exposure to the migrant education program. The fact that status, test type and grade level produce significant differences in student performance requires that subsequent analyses either control their interaction with entry age

or investigate that interaction directly. The effect of entry age on student performance within the other independent variables is the subject of EXPERIMENTAL SET B that follows.

B. EXPERIMENTAL SET B

Regression of scores at or above grade on entry age within status, test type, grade level.

1. Variables:

   Independent-Entry age, status, test type, grade level.
   Dependent-Percent of students scoring at or above grade level.

2. Population:

   General Effects data set. Each record contains a test type. For each test type, each record included a related: test outcome, status, entry age and grade level. Number of observations = 2,859. Number of cases = 351,011.

3. Tests:

   (Using GLM simple regression model): Regression of percent of test scores at or above grade:

   a. within each status,
   b. within each test type and
   c. within each grade level on entry age.

4. Hyp:Ho:

   For each analysis, slope of regression curve = 0.

   H1: For each analysis, slope not = 0.

5. Procedure:

   Three regression analyses were run--one for each independent variable to be controlled. The SAS, GLM, PROC statements used in each analysis are given below:

   (where PCT2=Percent of test scores at or above grade);

```
BY STATUS;              BY TEST_TYPE;          BY GRADE;
PCT2 = ENTRY_AGE;   PCT2 = ENTRY_AGE;    PCT2 = ENTRY_AGE;
```

   Each analysis was run against the General Effects data set without additional selection.

6. Results:

   For status 1, 2, 3 reject Ho that slope = 0. See Table B1 and Plot B1. For each test type reject Ho that slope = 0. See Table B2 and

Plot B2. For each grade level, reject Ho that slope = 0. See Table B3 and Plot B3. Archive SASTDGLM contains complete regression data and plots of actual and predicted values across entry age for each of the 12 grade levels. Table B3 is a synopsis of that data, and Plot B3 is a plot of the linear curve of best fit for each grade level.

7. Discussion of results:

From the analysis of variance of main effects (see Table A1), it has already been established that the main effect means are significantly different. That finding does not, however, tell which pairs of means within each main effect variable differ. Beyond the fact of significant differences among means of status, among means of test type and among means of grade level the regression analyses and plots allow us to make the following observations:

a. Inspection of Table B1 and Plot B1 indicates a high to low order of performance among the three status groups as follows: Status 3, 2, 1. Note that differences among the means are slight, and that no interactions are observed. Plot B1 also suggests that the relationship between status and entry age is linear.

b. Table B2 and Plot B2 indicate no interaction among test types. The curves seem to be linear (from a best fit perspective) and parallel. Of considerable program interest, however, is the grouping of curves. The two upper curves of Plot B2 are the arithmetic scores. The three lower curves are reading and language. The possibility exists that this general grouping by subject area may indicate test compatibility of the population being tested (e.g., arithmetic, being more symbolic, requires less English language facility to perform well during standardized testing in English than reading or language.) This grouping may alternately indicate population attributes that have little to do with testing (e.g., migrant students have been taught arithmetic in their native country but not English. Thus, the students do better in arithmetic because they are more skilled in it.) Whatever the phenomenon-suggested by the grouping of curves by subject area, it certainly appears worth further investigation. The possibilities are even more significant when one considers that the world of electronic information handling has strong symbolic content and may, therefore, be an area in which migrant students might enjoy success despite language handicaps.

c. Table B3 and Plot B3 show that, with the single exception of grade level 1, all regression curves have a negative slope. Thus, in all cases except grade level 1, student performance declines as entry age increases.

d. Plot B3 shows a primary and a possible secondary effect of entry age on student performance. The "primary" effect is the negative slope of each curve except that for grade level 1. The possibility of a secondary effect is discussed in "f" that follows. The meaning of the negative slopes may be paraphrased as follows: Performance on tests increases as entry age decreases. An even less formal statement would be that the earlier students enter the migrant program, the better they perform on tests. Thus entry age emerges as a legitimate determinant of student performance. A more direct investigation of the effect of entry age on performance is the subject of EXPERIMENTAL SETS D and F.

e. A loss of 44.37 percent occurs in moving from the mean (pct at or above grade) or grade level 1 to the mean (pct at or above grade) of grade level 12. If the entry age and grade level associated with each test were independent, then the gap between the curves would be attributable to increasing grade level. If, on the other hand, these two variables are not independent, then some of the gap may be attributable to entry age. This is the secondary effect of entry age referred to in "d" preceding. In fact, if entry age and grade level are (statistically) related, then identifying the amount of the gap(between the curves representing each grade level) to assign each variable becomes a complex undertaking. See "f" for a discussion of the possible relatedness of entry age and grade level.

f. It is possible that grade level and entry age are highly correlated. If so, two constraints on subsequent analyses must be observed. The first constraint is that grade level and entry age may not appear together as factors in a two (or more) factor analysis of variance. The second constraint is more subtle. If, in fact, these two factors are highly correlated, then care must be taken not to ascribe effects to one that are due to the other. Care must be exercised not to credit entry age with effects that materialized because entry age is highly correlated with grade level. To insure that further results are not "masked" effects of correlations, this study will next investigate the correlation between grade level and entry age (EXPERIMENTAL SET C). If the suspected correlation exists, then all subsequent analyses of the effect of entry age on test performance will be performed WITHIN grade level. (In SAS terminology, this means BY GRADE LEVEL.)

g. There is an apparent interaction between grade level and test performance. This interaction will be tested directly in EXPERIMENTAL SET E.

C. EXPERIMENTAL SET C

Relationship Between Entry Age and Grade Level.

1. Variables:

   Entry age, grade level.

2. Population:

   General Effects data set. Each record contains a test type. For each test type, each record included a related: Test outcome, status, entry age and grade level. Number of observations = 213. Number of cases = 351,011.

3. Test:

   Regression analysis using GLM simple regression model.

4. Hyp:Ho:

   The slope of the regression line of (mean entry age, mean grade level) = 0.

   H1:

   Slope not = 0.

5. Procedure:

   For each entry age, compute mean grade level. For each grade level, compute mean entry age. Run regression (of entry age on grade level and of grade level on entry age) analysis on means of independent variables using GLM simple regression model.

6. Results:

   Reject Ho for (grade level, entry age). Slope of regression lines of (entry age, grade level) are not equal to zero. See Table C1 and C2. Plots C1 an C2 show actual vs. predicted values for each regression pair (regression of mean entry age on mean grade level and vice versa).

7. Discussion of results:

   a. Inspection of the R-SQUARE values of Tables C1 and C2 show that entry age and grade level are highly correlated. This correlation means that the real effects on test performance of

entry age and grade level depend, to some degree, on the nature of their relationship. Definition of this relationship is the subject of the remainder of this experimental set.

b. Plots C3 and C4 are presented to illustrate one aspect of the relationship between entry age and grade level. Plot C3 shows the range of (90%) of entry ages associated with each grade level on the tests. Plot C4 shows the range of (90%) of the grade levels associated with each entry age on the tests. Notice that as grade level increases, the range of entry ages also increases (from Plot C3); whereas, as entry age increases, the range of grade levels decreases (from Plot C4). This effect is illustrated in Plot C5 as degrees of overlap of distributions of grade levels (see Plot C5-1) as entry age is varied from a value of 8 to 16 versus degrees of overlap of distributions of entry ages (see Plot C5-2) as grade level is varied from a value of 2 to 10.

Plots C5-1 and C5-2 seem to suggest that increased grade level (of test scores) might not add as much "grade level" effect to test performance degradation as increasing entry age (of test scores) might add "entry age" effect. Increasing from grade level 12 to grade level 10 added 56.75% of entry ages that were all greater than the maximum entry age in the grade level 2 distribution. Conversely, increasing from entry age 8 to entry age 16 added only 20.33% of grade levels that were higher than the maximum grade level in the entry age 8 distribution. This relative degree of effect would be true only if a "unit" of grade level had the same effect as a comparable unit of entry age. Plots C6 and C7 demonstrate the relative amounts of increase/non-increase in each variable as a function of increases in the other variable.

Plots C8 and C9 present the relationships being discussed in a more familiar manner. Notice in Plot C9 that:

*the means increase at the same rate,
*the standard deviation of entry ages increases at the same rate as does the mean entry age,
*the standard deviation of grade level decreases with increases in mean grade level.

The coefficients of variation in Plot C8 simply portray the relationship between the means and standard deviations more directly.

grade level for that student. This component of the rule has the effect of eliminating a zero (0) skills count for each record meeting the condition.

The rationale for use of the above rule is simply to eliminate from analysis those programs that probably had no skills reporting system in place at the time.

To insure that the presence or absence of zero (0) skills records did not bias the results unknowingly, each analysis was run in pairs--once including zero skills records and once excluding zero skills records.

D. EXPERIMENTAL SET D

Main Effects of Grade Level, Status and Test Type on Test Scores.

EXPERIMENTAL SET A will be performed again, but without the variable entry age being used as a factor. Entry age will be used instead to generate repeated measures so that interactions may be examined. Although entry age is not included formally as a variable in EXPERIMENTAL SET D, the fact that entry age and grade level are related means that some of the effects attributed to grade level are partially related to entry age.

1. Variables:

   Independent-Grade level, status, test type.
   Dependent-Percent of students scoring at or above grade level.

2. Population:

   General Effects data set. Each record contains a test type. For each test type, each record included a related: test outcome, status, entry age and grade level. Number of observations = 2,859. Number of cases = 355,011.

3. Test:

   Analysis of variance using GLM main effects model. A Duncan range test of means was also run.

4. Hyp:Ho:

   Main effect means within independent variables are equal.

   H1:

   Main effect means within independent variables are not equal.

5. Procedure:

   A GLM main effects analysis of variance with all interaction terms was run. Repeated measures were generated by using entry level as a dummy variable. Even though the N's per observation were unequal, a Duncan range test of means was included. The Duncan test proved to be sensitive to unequal N's but was useful where differences between means were extreme.

6. Results:

   Reject Ho for all main effect means and for all interaction terms. See Table D1.

7. Discussion of results:

a. The R-SQUARE value of 0.909914 (Table D1) indicates a good fit of the model. Even though this implies that much of the variability among the means is accounted for without entry being present as a factor, remember that correlation is a two-way street and some of the effect of grade level is because of its relationship with entry age (Entry age increases with increases in grade level.).

b. Though significantly different, the means for status were not sufficiently different to offset the unequal N's in order for the Duncan range test to identify the pairwise differences. See Table D2 for the means for status 1, 2 and 3.

c. The suspected grouping of test type scores by subject area is confirmed by the Duncan range test. CTBS arithmetic and CAT arithmetic means were not significantly different. CTBS reading, CTBS language and CAT reading means were not significantly different. However, the arithmetic means were significantly different from the reading/language means. Refer to the discussion of results "c" of EXPERIMENTAL SET B for further treatment of these findings.

d. The disparity of N's between grade levels makes the Duncan range test of means unreliable. Visual inspection of the means of Table D2 reveals a reversal of grades 6, 7 and 8 means. These means are very close in value however, and the reversal may simply be by chance. However, a legitimate subgrouping of means may be present and may be worth further investigation.

e. The interaction of status and test type follows the trend set by test type itself. Means for arithmetic are higher than means for reading/language regardless of status. Whether status enhances or mitigates this trend cannot be determined from the Duncan range test. More on this interaction appears in the next experimental set.

f. A discussion of the interaction of grade level with the other independent variables will be presented in the experimental set to follow.

E. EXPERIMENTAL SET E

Regression of Scores at or Above Grade on Grade Level Within; Status, Test Type and Entry Age.

1. Variables:

   Independent-Entry age, status, test type, grade level.
   Dependent-Percent of students scoring at or above grade level.

2. Population:

   General Effects data set. Each record contains a test type. For each test type, each record included a related test outcome, status, entry age and grade level. Number of observations = 2,859. Number of cases = 351,011.

3. Tests:

   (Using GLM simple regression model). Regression of percent of test scores at or above grade:

   a. within each status,
   b. within each test type and
   c. within each entry age
      on grade level.

4. Hyp:Ho:

   For each test, slope of regression curve = 0.

   Hl:

   For each test, slope not = 0.

5. Procedure:

Three regression analyses were run--one for each independent variable to be controlled. The SAS GLM PROC statements used in each analysis are given below:

(where PCT2=Percent of test scores at or above grade);

```
BY STATUS;                BY TEST_TYPE;          BY ENTRY AGE;
PCT2 = GRADE_LEVEL;   PCT2 = GRADE_LEVEL;  PCT2 = GRADE_LEVEL;
```

Each analysis was run against the General Effects data set without additional selection.

6. Results:

For status 1, 2, 3 reject Ho that slope = 0. See Table E1 and Plot E1.

For each test type reject Ho that slope = 0. See Table E2 and Plot E2.

For each entry age 0-11, reject Ho that slope = 0. For each entry age 12-19, fail to reject Ho that slope = 0. See Table E2 and Plot E3.

NOTE: Because of the density of data, Plot E3 was produced manually for even numbered entry ages only. See Table E3 for regression data for each entry age.

7. Discussion of results:

From the analysis of variance of main effects (see Table D1) it has been already established that the main effect means are significantly different. Beyond the fact of significant differences among means of status, among means of test type and among means of grade level the regression analyses and plots allow us to make the following observations regarding interactions:

a. For each status and for each test type, all curves have a negative slope. Thus, for each status and for each test type, performance declines as grade level increases. There are many possible reasons (and combinations thereof) for this decline related to grade level. An actual increase in test difficulty, an increased effect of language handicap, the telling effects of the life of the migrant, increased effect of test bias with grade level are but a few of many possible reasons for the consistently observed decline in performance.

b. Inspection of Plot E1 seems to indicate no interactions among the status 1, 2 and 3 groups. Plot E1 also suggests a linear relationship between status and grade level. Inspection of Plot B1 and Plot E1 suggests that slopes of status when plotted against entry age is greater than the slopes of status when plotted against grade level. No direct test of this possibility was run.

c. Inspection of Plot E2 reveals the same characteristic separation of test scores by the general subject area of the test as has been observed throughout this study without exception (i.e., The arithmetic scores are grouped higher than the reading/language scores.). The curves seem to be linear (from a best fit perspective) and parallel. The curves shown in Plots B2 and E2 show little difference in slope, linearity or

rank. This similarity makes the discussion of results "c" of EXPERIMENTAL SET B even more important. Whether plotted against entry age or against grade level and no matter the status, migrant students score at or above grade in arithmetic with a significantly greater frequency than they do in reading or language.

d. Plot E3 shows the decrease of scores for each of various entry ages from 0-11 as grade level increases. Two distinct groups are visible in that plot. Entry age groups in the range 0-11 show a decline roughly throughout all increases in grade level. Entry age groups 12-18 show no loss in performance as grade level increases. From this regression analysis, it may be observed that decreases in performance of students with entry ages greater than eleven cannot be attributed to the effect of grade level. EXPERIMENTAL SET B demonstrated that student performance does decrease beyond entry age 11 WITHIN GRADE LEVEL. This finding (see Table B3) coupled with the present findings (see Table E3) makes entry age--not grade level--the best determinant of performance for students having an entry age of 11 or greater.

e. A comparison of the R-SQUARE value of Tables B3 and E3 reflect the relationships between entry age and grade level that were identified in EXPERIMENTAL SET C.

f. Generally speaking, the 0-11 entry age groups perform higher then the 12-18 entry age groups throughout the range of comparable grade levels. This plot suggests that although early entry age students lose some of their advantage, they retain enough advantage, to continue to perform better than late entry age students.

g. Plot E3 suggests a way of making a direct comparison of the migrant students with minimal exposure to the program against migrant students with a constant (within each grade level) amount of exposure to the program. On Plot E3, notice the initial data points for entry ages 10, 12, 13, ... , 18. These initial data points are a product of the fact that a later entry age will likely be in a higher grade. For example, it is not likely that a student entering the program at age 14 would be placed in a grade lower than the 6th grade. The student has been attending school--he has just not been in the migrant program before!! Therefore, for each grade level, there must be a group of students who are just entering the migrant program for the first time. These groups would constitute a "naive" control group. Such a design was devised and carried out as described in EXPERIMENTAL SET F.

F. EXPERIMENTAL SET F

A Comparison, at Each Grade Level, of Migrant Students Having Minimal Exposure to the Migrant Program With Migrant Students Who Entered the Migrant Program at Age Six.

NOTE: The term, experimental group, may be abbreviated as Expgr. The term, test type, may be abbreviated as TSCODE.

1. Variables:

   Independent-Expgr, grade level, test type, status.
   Dependent-Percent of group scoring at or above grade level.

2. Population:

   Two populations were selected from the General Effects data st. Experimental group, Expgr6, consisted of test scores of migrant students with an entry age of six. Experimental group, Expgr1, consisted of test scores of migrant students who, six months prior to being tested, had not ever been in any migrant program. See Table F1 for a review of factors and levels used in the analysis of variance. See Table F4 for a review of the observations and cases used in the regression analysis.

3. Test:

   The first test run was a GLM analysis of variance of main effect (independent) variables and of all interaction terms. The second test was the regression of mean test scores of each experimental group on grade level. A third test (GLM analysis of variance) was run to check for a difference in age between the two experimental groups.

4. Hyp:Hoa:

   Means of main effect (independent) variables are equal.

   Hob:

   The slope of the regression curve of each experimental group = 0.

   Hoc:

   The two groups do not differ with respect to age.

   Hla:

   Means of main effect variables are not equal.

H1b:

The slope of the regression curve of each experimental group is not = 0.

H1c:

The groups do differ with respect to age.

5. Procedure:

The two experimental groups were formed. The tests identified above were then run.

6. Results:

Reject Hoa. Table F2 shows the results of the analysis of variance and Table F3 displays the means of the main effect variables.

Fail to reject Hob. Table F5 shows the means, by grade level, for each experimental group. Both slopes of the regression curves differed from 0 as may be observed in Tables F6 and Plot F1.

Fail to reject Hoc. Tables F7 show the results of the analysis of age.

7. Discussion of results:

a. The means of the experimental groups differ overall and by grade level, test type, status and various combinations thereof as Table F2 demonstrates. The fact that the means of all independent variables showed significant differences indicates that the effect of these variables is the same as it has been in previous analysis (see EXPERIMENTAL SET D and E). The interactions not involving the experimental groups do differ from prior analyses. Test type X status is no longer a significant interaction nor is the three way interaction, grade level X, test type X status. Thus in most ways, the experimental groups resemble the parent data set.

b. Of particular interest is the effect of grade level on the two experimental groups. Plot F1 shows this effect. The group with entry age of six is superior at each grade level. Both groups' performance declines with increasing grade level but, the entry age six group declines less than the minimal exposure group over the range of grade levels.

c. It was suggested that in addition to the entry age six students performing better, that group might also be younger. In effect this would mean that not only does prior migrant program exposure produce higher performance, it (exposure) produces said higher performance at an earlier age thus conserving educational resources. An analysis of variance showed no significant difference of the mean ages of the two groups nor interaction between experimental group and grade level. See Table F6.

G. EXPERIMENTAL SET G

Joint effect of entry age and grade level on student performance.

EXPERIMENTAL SET A examined the main effect of status, test type, entry age and grade level on student performance. During that examination the possibility of a relationship between entry age and grade level emerged.

EXPERIMENTAL SET B examined the effect of entry age on student performance while holding all other independent variables constant.

EXPERIMENTAL SET C examined the relationship between entry age and grade level.

EXPERIMENTAL SET D examined the main effects of grade level, status and test type (with interactions) on student performance.

EXPERIMENTAL SET E examined the effect of grade level on student performance while holding other independent variables constant.

EXPERIMENTAL SET F examined the effect of entry age on a control versus an experimental group.

EXPERIMENTAL SET G concludes the study of the effects of entry age and grade level by examining their joint effect on student performance.

1. Variables:

   Independent-Entry age, grade level.
   Dependent-Percent of scores at or above grade level.

2. Population:

   General Effects data set. Each record contains a test type. For each test type, each record included a related: Test outcome, status, entry age and grade level. Number of cases - 351,011.

3. Tests:

   Four tests were run. Each "test" is, in fact, a model which uses the Least Squares Method of fitting the data.

   Model 1: Test 1 was a simple multiple (regression) analysis of the regression of student performance on entry age and grade level.

   Model 2: Test 2 was a polynomial multiple (regression) analysis of the regression of student performance on entry age and grade level and entry age X grade level.

Model 3: Test 3 was a full-model polynomial multiple (regression) analysis of the regression of student performance on entry age and grade level and grade level X grade level and entry age X entry age and entry age X grade level.

Model 4: Test 4 was constructed on the basis of observing the outcomes of Tests 1-3 above. Specifically, the significant terms of Test 3 were chosen for the Test 4 model. Test 4 was a (partial model) polynomial multiple (regression) analysis of the regression of student performance on grade level and grade level X grade level and entry age X entry age.

The use of multiple regression analysis yields predictor models consisting of surfaces rather than simple curves.

4. Hyp:Ho:

All null hypotheses were to the effect that all deviations of regression terms from 0 were due to chance.

Hl:

All alternative hypotheses attributed deviations of regression terms from 0 to other than chance.

5. Procedure:

The SAS, GLM, PROC statements for each test is given below:

Model 1:PCT2=ENTR_AGE GRADE
Model 2:PCT2=ENTR_AGE GRADE ENTR_AGE*GRADE
Model 3:PCT2=ENTR_AGE GRADE ENTR_AGE*ENTR_AGE GRADE*GRADE
Model 4:PCT2=GRADE GRADE*GRADE ENTR_AGE*ENTR_AGE
(Each model included a WEIGHT statement).

Where

PCT2 means percent of test scores (regardless of test type) at or above grade level.

ENTR_AGE is the linear component for entry age.

GRADE is the linear component for grade level.

ENTR_AGE*ENTR_AGE is the quadratic component for entry age.

GRADE*GRADE is the quadratic component for entry age.

GRADE*GRADE is the quadratic component for grade level.

ENTR_AGE*GRADE is the interactive term for the interaction of entry age and grade level.

6. Results:

Table G1 shows that when a linear model is chosen, both entry age and grade level are significant.

Table G2 shows that Model 2 has a slightly higher R-SQUARE than Model 1. This slightly better fit is probably because of the addition of the interactive term (entry age X grade level) to the model.

Table 3 shows the results of the full-model. Neither the linear entry age term nor the entry age X grade level interactive term is significant.

Table 4 shows the results of deleting from the analysis the terms of Model 3 that were not significant. There is very little difference in R-SQUARE between Model 4 and Model 3. Thus, Model 4 accounts for as much of the variability as does Model 3 but with two less terms.

7. Discussion of results:

Model 4 has been selected (from among the four models) as best describing the relationship between entry age and grade level. This model has a higher R-SQUARE value than Model 1 or 2 and fewer terms than Model 3.

In Table 4, the opposite signs of the two interactive terms (ENTR_AGE*ENTR_AGE and GRADE*GRADE) indicates a <u>hyperbolic</u> relationship. This coupled with the parabolic nature of each interactive term yields a surface described as a <u>hyperbolic paraboloid</u> (visually resembles the seating surface of a saddle).

Plot G5 is a display of the joint effect of two independent variables (entry age and grade level) on student performance. This representation is constructed by plotting the unweighted percents against each combination of entry age and grade level for which there was data. The term, unweighted, means that the frequency with which each score (percent at or above grade) occurs for each of the combinations of independent variables is not reflected in the plot.

Inspection of Table G4 and Plot G1 permits the following observations:

a. Although Plot G1 is unweighted, it presents a surface which is not unlike that identified by Model 4. Plot G1 displays key features identified by Model 4 as indicated in "b" and "c" below.

b. The GRADE*GRADE quadratic term, being positive in sign, implies a parabola which opens upward. Thus, that portion of the parabolic curve which corresponds to real data describes an effect which diminishes across its range. This is readily seen in Plot G1 by observing that the amount of drop in performance decreases as grade level increases.

c. The ENTRY_AGE*ENTRY_AGE quadratic term, being negative in sign, implies a parabola which opens downward. Thus, that portion of the parabolic curve which corresponds to real data describes an effect which increases across its range. This is readily seen in Plot G1 by observing that the amount of drop in performance increases as entry age increases.

d. The hyperbolic character of the data, as identified by the model, is not readily apparent from the data.

In summary, the relationship between entry age and grade level and the joint effect of these variables on student performance is a complex one. The response surface identified by Model 4 provides a useful map when dealing with the effects of either of these variables on student performance.

TABLE OF CONTENTS

PART III

EXPERIMENTAL SETS H-K

EXAMINATION OF THE MATH SKILLS SYSTEM
AND
ITS RELATIONSHIP TO STUDENT PERFORMANCE

Introduction to the Skills Data.

An important note on calculating the mean number of skills per student.

		PAGE
H.	EXPERIMENTAL SET H	III-1
	Descriptive statistics of math skills data in the SKILLS data set.	
I.	EXPERIMENTAL SET I	III-7
	Descriptive statistics of math topics represented by the math skills data in the SKILLS data set.	
J.	EXPERIMENTAL SET J	III-10
	Descriptive statistics of math skills data and math test data in the SKILLS/TEST data set.	
K.	EXPERIMENTAL SET K	III-26
	Migrant student performance on standardized math tests as a function of varying numbers of math skills in MSRTS records.	

Note 1: When inspecting the analysis of variance tables presented herein, always refer to the TYPE III values (as opposed to TYPE I).

Note 2: Most of the data plots represent very large samples relative to the population. Vertical differences between data points may often appear not to be significant when in fact they are.

Note 3: Simple regressions are normally run for A on B and B on A. For this reason, the independent/dependent variables involved in the regression analyses are simply specified as members of the set (A, B).

Note 4: All testing is done at alpha = .05.

****
* All Tables and Plots referenced herein will be found in the *
* accompanying ILLUSTRATIONS MANUAL. *
****

Introduction to the Skills Data

Part II preceding examined some of the overall relationships between the migrant program and student performance on standardized tests. Part III will examine the math skills system data and the relationship between that data and student performance on standardized tests. Detailed information regarding the formation of the data sets used in Part III may be found in the Appendix, Section C, Populations Selected for Use in Studying the Effects of the Skills System on Standardized Scores in Math and Reading. A summary of the numbers of skills and/or tests making up the populations studied will be presented in each experimental set.

An Important Notice on Calculating the Mean Number of Skills Per Student.

When calculating the mean number of skills (by grade level) for each student, there is a fundamental decision to be made that affects the resulting values greatly. That decision is whether to include cases in which students have no (0) skills reported at a given grade level. The inclusion of the numeric value zero (0) will of course have the effect of producing lower means because it raises the number of cases (the denominator) without changing the total skills counted (the numerator). Because of the non-uniform implementation schedule of the skills reporting system across the national program, the lead author decided that the inclusion of ALL cases of no skills would not be meaningful and would introduce uncertainty as to the representativeness of data sets. The author did not want to mix data from programs lacking skills reporting with data from programs having implemented skills reporting. In addition, the author wished to avoid the possibility of unknowingly introducing underlying bimodal distributions of skills (i.e., frequency distributions of skills in which a non-trivial portion of the population has zero skills and the remainder has a more normal spread of skills). To exclude ALL cases of zero (0) skills reported would misrepresent the realities of the program in the opposite direction. A compromise solution consisting of the following rule was adopted:

> If reading skills but no math skills were reported for a student (for a given grade level) then a math skills record was created and assigned the value of zero (0) for the number of math skills reported for that student at that grade level.
>
> If neither reading nor math skills were reported for a student at a given grade level then no math skill record was created at that

grade level for that student. This component of the rule has the effect of eliminating a zero (0) skills count for each record meeting the condition.

The rationale for use of the above rule is simply to eliminate from analysis those programs that probably had no skills reporting system in place at the time.

To insure that the presence or absence of zero (0) skills records did not bias the results unknowingly, each analysis was run in pairs--once including zero skills records and once excluding zero skills records.

H. EXPERIMENTAL SET H

Descriptive Statistics of Math Skills Data in the SKILLS Data Set.

SKILLS Data Set Description

The SKILLS data set represents all cases in which a student had at least one math or reading skill reported under a grade level 1-12 enrollment.

EXPERIMENTAL SET H will examine the math skills present in three populations drawn from the SKILLS data set. These populations are specified under item 2. Populations, below.

1. Variables:

   Distribution statistics BY GRADE LEVEL including mean number of skills per student for each of the above following populations.

2. Populations:

   1. Those records containing at least one math skill
   2. Those records containing at least one math skill or at least one reading skill or both
   3. Those records containing at least one math skill and at least one reading skill.

3. Test: For each population:

   (Using GLM analysis of variance): Test of difference among mean number of math skills at each grade level.

   (Using GLM simple regression model): Regression of mean number of math skills on grade level.

   (Using GLM polynomial regression model): Regression of mean number of math skills on grade level.

4. Hyp:Hoa:

   Within each population, there is no significant difference among the mean numbers of math skills per student at different grade levels.

   Hlb:

   Difference among means.

   Hob:

For each population, the slope of simple regression of mean number of math skills on grade level = 0.

H1b:

Slope not = 0.

Hoc1:

For each population, the linear term for the regression of the mean number of math skills on grade level = 0.

Hoc2:

For each population, the quadratic term for the regression of the mean number of math skills on grade level = 0.

H1c1:

Linear term not = 0.

H1c2:

Quadratic term not = 0.

5. Procedure:

*Distribution statistics for each population were reported.
*Tests as specified were run.

6. Results:

See Table H1 for a summary of the analyses performed on all three populations.

Population 1: (Each record contained at least one math skill.)
(See Tables H5, H6 and H7.)

Significant

Difference of mean number of skills among grade levels.

Slope of regression of mean number of skills on grade level.

Not significant

Linear and quadratic term of (polynomial) regression of mean number of math skills on grade level.

Population 2: (Each record contained at least one math skill or at least one reading skill or both.) (See Tables H8, H9 and H10.)

Significant

Difference of mean number of skills among grade levels.

Linear term of polynomial regression of mean number of math skills on grade level.

Quadratic term of polynomial regression of mean number of math skills on grade level.

Not Significant

Slope of regression of mean number of skills on grade level.

Population 3: (Each record contained at least one math skill and at least one reading skill.) (See Tables H11, H12 and H13.)

Significant

Difference of mean number of skills among grade levels.

Slope of regression of mean number of skills on grade level.

Not Significant

Linear and quadratic term of polynomial regression of mean number of math skills on grade level.

7. Discussion of Results:

Population 1: (Each record contained at least one math skill.)

i. There is a significant but slight positive linear relationship between the mean number of math skills per student and grade level (Table H1, H2, H5, H6 and Plot H1).

ii. The low R-SQUARE value of the analysis of variance indicates that some other independent variable (e.g., entry age) may be a stronger determinant of variability.

iii. Neither the linear nor the quadratic term of the polynomial regression was significant (Table H7).

In view of the low R-SQUARE values for the ANOVA and for the linear regression, it may be observed, conservatively, that for this population, the mean number of math skills per student does NOT decrease by grade level. This observation is corroborated by the means presented in Table H2.

For this population and population 3, neither the analysis of variance model nor the regression model accounts for much of the variability in the mean number of skills per student (refer to the R-SQUARE values in Table H1). These low values indicate that grade level is not a strong determinant of variability among mean numbers of skills per student for these two populations. The factor most likely affecting the variability among the means is entry age. In Part II of this report, it was observed that entry age increases with grade level. Later entry ages means fewer cumulative skills in all student record and fewer current skills in student records in the higher grade levels (there are fewer math skills appropriate to the higher grade levels). It is possible that an analysis of mean skills per student which accounts for differing entry ages within each grade level would yield very high (i.e., $>$.90) R-SQUARE values.

Population 2: (Each record contained at least one math skill or at least one reading skill or both.)

i. Although the mean number of skills per student does differ across grade level (Tables H1 and H8), these differences constitute neither a positive nor a negative linear regression (Tables H1, H3 and H9).

ii. The best description of this population is provided by the polynomial regression model which shows an R-SQUARE of .7855 (Tables H1 and H10 and Plot H2). The curve representing this population is a positively accelerated curve (Plot H2).

iii. The means for grade levels 9-12 seem to form a subgroup that has substantially fewer skills (Table H2 and Plot H2).

For the general population of migrant students attending schools demonstrating the presence of a skills reporting system, the mean number of math skills reported per student increases until about the sixth grade and then decreases through the twelfth grade.

Population 3: (Each record contained at least one math skill and at least one reading skill.)

i. There is a significant but slight positive linear relationship between the mean number of math skills per student and grade level (Tables H1, H4, H11, H12 and Plot H3).

ii. The low R-SQUARE value of the analysis of variance indicates that some other independent variable (e.g., entry age) may be a stronger determinant of variability.

iii. Neither the linear nor the quadratic term of the polynomial regression was significant (Table H13).

In view of the low R-SQUARE values for the ANOVA and for the linear regression, it may be observed, conservatively, that for this population, the mean number of math skills per student does NOT decrease by grade level. This observation is corroborated by the means presented in Table H4.

This population behaves in a similar fashion to population 1 which is not too surprising since population 3 is a subset of population 1 (i.e., Each record containing a math and a reading skill contains a math skill.).

General Discussion of Results:

In essence, these results may be rephrased as follows:

For those students on whom math skills are reported (population 1), the mean number of math skills reported does not decrease with grade level.

For students overall (population 2), the mean number of math skills reported decreases with grade level--especially in grades 9-12.

The population at large has 406,559 records (Table H3). This number represents the maximum number of records that may contain at least one math skill. Table H2 shows that 337,478 records do indeed contain at least one math skill record. Thus, it may be stated that about 83% (337,478 divided by 406,559) of all possible SKILL records contain one or more math skills. Taking this analysis a step further, it may be calculated that of the 59,478 first grade records, 84.9% (50,534 divided by 59,478) possess math skills data.

By similar reasoning, the percent of total possible cases containing math skills may be computed at each grade level. A summary of such percents appears below:

Percent of Total Possible Cases Containing Math Skills at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
84.9	85.2	86.0	85.6	85.6	84.4	79.9	78.2	64.2	56.7	47.5	39.8

The above percents may be viewed as the percent of students on whom math skills are reported at each grade level. Note that the percent decreases sharply in grades 9-12.

The preceding calculations and the results being reviewed lead to two observations of interest:

(Observation 1)

****
* The percent of enrolled students on whom math skills are reported *
* decreases with increases in grade level. The amount of decrease *
* is greatest for grades 9-12. *
****

(Observation 2)

****
* For those students having math skills reported, the mean number *
* of skills reported per student increases slightly across grade *
* levels. *
****

Two important questions arise from these two preceding observations:

What are the attributes of those students on whom skills continue to be reported beyond grade level 9?

Which skills are being reported in the higher grade levels?

The above questions will be addressed in EXPERIMENTAL SET J.

I. EXPERIMENTAL SET I

Descriptive Statistics of Math Topics Represented by the Math Skills Data in the SKILLS Data Set.

SKILLS Data Set Description

The SKILLS data set represents all cases in which a student had at least one math or reading skill reported under a grade level 1-12 enrollment.

1. Variables:

   The frequency of math skills occurring under each math topic by grade level.

2. Populations:

   Those records in the SKILLS data set containing at least one math skill.

3. Tests:

   None.

4. Hypotheses:

   None.

5. Procedure:

   A grade level by math topic frequency table of math skills was prepared for the population.

6. Results:

   See Table I1 in the ILLUSTRATIONS MANUAL.

7. Discussion of results:

   One of the important observations to be drawn from Table I1 is that there are fewer math skills in the skills system that apply to the higher grades as there are skills which apply to the lower and mid-level grades. The figure following shows the grade level at which the greatest percent of skills within each math topic area are reported.

Topical grade level at which greatest area percent of skills are reported	
Readiness	1
Number Meaning	1
Whole Numbers	6
Fractions	6
Decimals	7
Percent	7
Measurement	2
Geometry	8
Stat/Prob	8
Sets	1

Judging from usage in the field, there are few math skills that are appropriate beyond the ninth grade level. This circumstance is being remedied by a new version of the skills system.

The table below shows the top five math topics reported at grade levels 9-12. This table also shows the percent of total skills reported at each grade level that each topical area represents. For example, 25.09% of all skills reported at the ninth grade level were in the topical area of FRACTIONS. Of all math skills reported at the ninth grade level, 91.49% fall under the five topical areas shown on the table.

Grade Level---->	9	10	11	12
TOPIC				
Fractions	25.09	25.67	27.93	30.97
Whole Numbers	22.04	17.56	15.45	16.50
Percent	15.55	17.19	16.00	13.67
Decimals	14.86	13.64	12.88	12.80
Measurements	13.95	15.11	15.89	13.50
Total %	91.49	89.17	88.15	87.44

Previous observations and the data of Table H10 suggest the following conclusions:

From EXPERIMENTAL SET I:

1. The skills being reported at grade levels 9-12 are:
   inappropriate to those grade levels in the "normal" course of instruction, and
   reflect instruction aimed at teaching skills that students are intended to acquire at earlier grade levels.

From EXPERIMENTAL SET H:

ii. The percent of all students on whom math skills are reported decreases sharply in grade levels 9-12.

iii. For those students on whom math skills ARE reported in grade levels 9-12, the mean number of math skills reported per student does not decrease as grade level increases from grade level 1-12.

From the above conclusions the following summary statement may be made:

****
* The skills data reported for grade 9-12 migrant students *
* represents a very specific subset of (9-12) students for *
* whom instruction is conducted that is aimed at teaching *
* those math skills that are ordinarily acquired at earlier *
* grade levels. *
****

J. EXPERIMENTAL SET J

Descriptive Statistics of Math Skills Data and Math Test Data in IDSHTDSK. For Clarity, the Data Set IDSHTDSK Will be Called "the SKILLS/TEST Data Set."

SKILLS/TEST Data Set Description

The SKILLS/TEST data set represents all cases in which a student had math or reading test data reported under a grade level 1-12 enrollment. Table J1 in the ILLUSTRATIONS MANUAL gives essential parameters of the SKILLS/TEST data set and its derivation from the MSRTS data base. The reader is referred to the Appendix, Section C for details of the SKILLS/TEST data set (identified as IDSHTDSK in the Appendix) and Appendix, Section C for a definition of "current + prior" and "cumulative" skills.

A total of eight populations were drawn from the SKILLS/TEST data set and analyzed with respect to both the mean number of math skills per student and the mean number of math tests per student. In a very important sense, this is the world of skills and tests that the world looks at in standardized testing. These populations are not equally important to the test at hand and will be discussed accordingly.

1. Variables:

   (DEPENDENT VARIABLE = INDEPENDENT VARIABLE)
   MEAN NUMBER OF MATH SKILLS PER STUDENT = GRADE LEVEL
   MEAN NUMBER OF MATH TEST PER STUDENT = GRADE LEVEL
   MEAN NUMBER OF MATH TEST PER STUDENT = MATH SKILLS/STUDENT

   The above three relationships will be examined for each of the eight populations listed below. In addition, the relation, mean math skills per student=grade level will be examined "current + prior" and "cumulative" skills for populations 4 and 7.

2. Populations:

   1. Those records containing at least one math skill.

   2. Those records containing at least one math skill or at least one reading skill or both.

   3. Those records containing at least one math skill and at least one reading skill.

   4. Those records containing at least one math test.

5. Those records containing at least one math test or at least one reading test or both.

6. Those records containing at least one math test and at least one reading test.

7. Those records containing at least one math skill and at least one math test.

8. Those records containing at least one math skill or at least one reading skill and at least one math test.

3. Tests:

For each population:

(Using GLM analysis of variance): Test of difference among mean number of math skills at each grade level.

(Using GLM analysis of variance): Test of difference among mean number of math tests at each grade level.

(Using GLM analysis of variance): Test of difference among mean number of math tests as a function of varying numbers of math skills in student records.

(Using GLM simple regression model): Regression of mean number of math skills on grade level.

(Using GLM simple regression model): Regression of mean number of math tests on grade level.

(Using GLM simple regression model): Regression of mean number of math tests on varying numbers of math skills in student records.

(Using GLM polynomial regression model): Regression of mean number of math skills on grade level.

(Using GLM polynomial regression model): Regression of mean number of math tests on grade level.

(Using GLM polynomial regression model): Regression of mean number of math tests on varying numbers of math skills in student records.

4. Hyp:Hoa:

Within each population, there is no significant difference among the mean numbers of math skills (or tests) per student at different grade levels (or numbers of skills).

H1b:

Difference among means.

Hob:

For each population, the slope of simple regression of mean number of math skills (or tests) on grade level (or numbers of skills) = 0.

H1b:

Slope not = 0.

Hoc1:

For each population, the linear term for the regression of the mean number of math skills (or tests) on grade level (or numbers of skills = 0.

Hoc2:

For each population, the quadratic term for the regression of the mean number of math skills (or tests) on grade level (or number of skills) = 0.

H1c1:

Linear term not = 0.

H1c2:

Quadratic term not = 0.

5. Procedure:

*Distribution statistics for each population were prepared.
*Tests as specified were run.

6. Results:

See Tables J2, J3 and J4 for summaries of the analyses performed on all eight populations.

The detailed tables of results for all analyses over all populations will not be presented because they are too numerous (3 pair of dependent/independent variables X 3 types of analysis X 8 populations = 72 tables). These tables as well as all other data, tables and charts referenced directly or indirectly in this study have been catalogued and are on file for the authorized readers inspection at the MSRTS in Little Rock, Arkansas.

7. Discussion of Results:

VARIABLES: MEAN NUMBER OF MATH SKILLS PER STUDENT = GRADE LEVEL

(Type of analysis: ANOVA)

Notice from Table J2 that all differences between mean numbers of skills per student across grade levels were significant. However, the actual differences varied enormously depending on the population--but with a pattern.

The chart below shows, for each population;

the highest mean number of math skills reported in any grade level, the lowest mean number of math skills reported in any grade level, and the percent of change in means.

NOTE: The grade level at which the highest/lowest mean occurred is in parentheses.

Present in each record making up the population	HIGH MEAN	LOW MEAN	PERCENT OF CHANGE
1 math skills	9.45 (7)	7.39 (9)	21.79
2 math skills or read skills	7.87 (1)	2.46 (12)	68.74
3 math skills and read skills	9.93 (8)	7.98 (9)	19.73
4 math test	5.64 (1)	0.28 (12)	95.03
5 math test or read test	5.49 (1)	0.27 (12)	95.08
6 math test and read test	5.61 (6)	0.26 (12)	95.36
7 math skill and math test	9.65 (8)	7.36 (9)	23.73
8 math skill or read skill and math test	8.20 (5)	2.35 (12)	71.34

The following observations are in order regarding Table J2 and the above chart of means:

i. For all populations, the very large N's (large number of cases) within grade level produce a relatively small mean square error term which in turn produces a significant F ratio. This is enhanced by the fact that the degrees of freedom for the "model" are always low at 11 (i.e., total number of grade levels -1). This is not to imply that the difference among means are not "real." They certainly are!

ii. The extremely low R-SQUARE values associated with these results indicate that the ANOVA model provides a very low accounting of the variability within grade level.

iii. From the preceding chart it may be observed that the magnitudes of differences among population high/low means fall into two groups. Populations 1, 3 and 7 show less than 24% difference between the high and low means, and the remaining populations show 68% or greater difference. The separation of differences is accounted for by the fact that in each record of population 1, 3 and 7, at least one math skill must be present. This condition is not true for the other populations.

(Type of analysis: REGRESSION)

i. Although all populations show significant differences among the mean number of math skills per student across grade levels, no population in which math skills were present in each record (by definition) showed a regression of any direction or degree on grade level. This result is in keeping with the findings of EXPERIMENTAL SET H regarding the SKILLS data set--namely, for those students on whom math skills are reported, the mean number of math skills reported remains constant across grade levels 1-12.

ii. All of the five populations in which math skills need not be present (by definition) show a significant negative second order regression on grade level. The plots for these five curves are shown on Plot J1 as predicted values. The POLYNOMIAL R-SQUARE values in Table J2 show a good fit for these populations. These results are also in agreement with the SKILLS data set which showed that when considering all migrant students attending schools where skills reporting is in operation, the mean number of math skills reported per student decreases over grades 1-12 and especially over grades 9-12.

iii. Math skills data is distributed over the SKILLS data set and over the SKILLS/TEST data set in a similar fashion. This conclusion is based on the following two findings which are true for both data sets:

Populations composed of records which by definition always contain math skills show no regression on grade level.

Populations composed of records which by definition may not always contain math skills all show a negative second degree polynomial regression on grade level as the best fit model for describing the data.

iv. The finding that math skills data is distributed over both data sets in a similar fashion does not mean that the distribution of math skills is the same over all migrant students as it is over all migrant students who are tested in math. This is an entirely different issue that will be dealt with briefly now and in-depth in EXPERIMENTAL SET K.

v. Population 4 in Table J2 represents the skill records of migrant students who are tested in math. Each record contains the results of one or more math tests and may or may not contain math skills data. It has been shown (see Table J2 and Plot J1) that the mean number of math skills per student in this population decreases across grade levels 1-12 and sharply across grade levels 9-12. This population represents the migrant student standardized math test scores that normally would be used to represent the entire migrant population in any kind of comparison with other groups (e.g., regular students, students in other compensatory programs, etc.).

Population 7 in Table J2 represents the skill records of migrant students who, at the same grade levels, have both math test scores and math skills data in their records. It has been shown (see Table J2 and Plot J1) that the mean number of math skills per student in this population does not decrease across grade levels 1-12. Since any record in this population contains math skills data, this collection of records is distributed in a similar fashion to Population 1 (see Table J2) and to the equivalent population, records containing math skills, in the SKILLS data set. EXPERIMENTAL SET H has shown that the number of students on whom math skills are reported declines sharply from grade levels 9-12. EXPERIMENTAL SET H also showed that for those students on whom math skills continue to be reported in grade levels 1-12, the mean number of skills reported from grade levels 1-12 does not decrease. A parallel analysis was performed on the SKILLS/TEST data set and is presented below.

vi. The SKILLS/TEST population at large (Population 5: records having math test data or reading test data) has 144,148 records. This number represents the maximum number of records that may contain math skills data AND math test data. Appendix Figure E-4 in the ILLUSTRATIONS MANUAL shows that 66,734 records do indeed contain both math skills and math test data. Thus, it may be stated that about 46% (66,734 divided by 144,148) of all possible SKILLS/TEST records contain both math skills and math test data. By similar reasoning, the percent of total possible cases containing both math skills and math test data may be computed at each grade level. A summary of such percents appears below:

Percent of Total Possible Cases on the SKILLS/TEST Data Set Containing both Math Skills and Math Test Data at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
55.1	55.7	56.0	56.7	56.3	52.2	42.2	39.6	13.3	7.0	3.9	2.6

The above percents may be viewed as the percent of students on whom math skills and math tests are reported at each grade level. Note

that the percent decreases sharply in grades 9-12. This percentage decrease could come from either of these sources:
a decrease (over grade levels) in the number of records containing math skills data, or

a decrease (over grade levels) in the number of records containing math test data, or

a change (over grade level) in the joint distribution of tests and skills.

The actual source of the percentage decrease (over grade level) in the number of records containing both math skills and math test data will be examined in paragraphs vii, viii and ix following:

vii. Based on the same rationale used in paragraph vi preceding, the percent of all possible cases in the SKILLS/TEST data set containing math skills was computed at each grade level and appears below:

Percent of Total Possible Cases in the SKILLS/TEST Data Set Containing Math Tests at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
60.7	59.9	59.4	60.3	59.7	55.2	45.6	45.6	14.8	8.5	4.8	3.1

It may be observed that the drop in the percent of cases containing math skills roughly parallels the drop in the number of cases having math skills and math tests in the percent summary appearing in paragraph vi preceding.

viii. Based on the same rationale used in paragraph vi preceding, the percent of all possible cases in the SKILLS/TEST data set containing math tests was computed at each grade level and appears below:

Percent of Total Possible Cases in the SKILLS/TEST Data Set Containing Math Tests at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
89.7	92.6	92.7	92.2	92.7	93.3	92.6	93.0	88.1	85.9	84.0	82.5

It may be observed that the drop in the percent of cases containing math tests is not characteristic of the drop in the number of cases having math skills and math tests in the percent summary appearing in paragraph vi preceding.

ix. If two variables are (statistically) independent, then their joint occurrence is given by the product of their individual occurrences. Based on this principle, the probability of joint

occurrence (expressed as a percent) of math skills and math tests, as two independent variables, was calculated at each grade level and is presented below. Beneath each calculated estimate of joint occurrence is the percent that the estimate deviates from the actual data as presented in paragraph vi preceding.

Estimate of the Percent of Total Possible Cases in the SKILLS/TEST Data Set That Would Contain Both Math Skills and Math Tests at Each Grade Level if Math Skills and Math Tests were Distributed Independently

1	2	3	4	5	6	7	8	9	10	11	12
54.4	55.4	55.0	58.9	55.3	51.5	42.2	42.4	13.0	7.3	4.0	2.5
1.2	0.5	1.7	3.7	1.7	1.3	0.0	6.6	2.2	4.1	2.5	3.8

Discounting rounding error (all values were truncated) two conclusions seem to be in order:

One: Math skills and math tests seem to be independently distributed over the SKILLS/TEST data set.

Two: If skills and tests are not independently distributed, then no pattern of non-independence seems to emerge. The greater percentage of discrepancies in the higher grades is probably related to sampling error induced by the smaller number of cases at those grade levels.

In effect, the above chart eliminates non-independence between the math skills and math tests as a source of the drop in the joint occurrence of math skills and math tests in grade levels 9-12.

x. Conclusions regarding the math skills and math tests in the SKILLS/TEST data set.

There is a sharp decrease in Grade levels 9-12 in the number of records having both math skills and math tests data.

This drop is most likely accounted for by a corresponding decrease in the number of Grade level 1-12 records containing math skills.

- -

All analyses of mean number of math skills as a function of grade level up to this point have been directed at "current" skills; that is, the math skills reported for the enrollments occurring at each grade level. Two additional measures of skills were derived and will be examined for Population 4 (all records having math tests) and Population 7 (all records having both math skills and math tests. These derived measures are "current + prior" math skills and "cumulative" math skills. These skills

are as their name suggests. For a detailed description of their derivation, see the Appendix, Section C, step 1 under "the derivation of IDSHTDSK (skills and test data)."

The curve of best fit for "current + prior" math skills and for "cumulative" math skills over Populations 4 and 7 were polynomials. These curves are shown in Plot J2 which, for comparison's sake, also included best fit curves for "current" skills. A discussion of the interpretation and implications of these curves follows:

There are certain properties of the distribution of cumulative math skills that hold for both Populations 4 and 7. The bottom three curves of Plot J2 show the distribution of the mean number of math skills per student across grade levels for Population 4 (records having math tests but not necessarily math skills).

The top three curves of Plot J2 show the distribution of the mean number of math skills per student across grade levels for Population 7 (records having both math tests and math skills).

Note that the two curves displaying the cumulative skills (one for each population) run counter to the common sense notion that a cumulative quantity cannot decrease. The decrease in the cumulative skills as grade level increases beyond the fifth grade is a result of three factors which make the analysis of migrant student data different from that of regular students.

The first factor is the recognition that the math skills being discussed are a part of the migrant education program rather than the regular school program. Thus, skills accrue to a student's record only if that student is enrolled in a migrant education program.

The second factor is that of entry age--the age of the migrant student when first enrolled in a migrant education program. Entry age increases significantly with grade level. This means that migrant students in the higher grade levels have been members of migrant education programs for a far shorter period of time than have migrant students in the mid-level grades. Thus, migrant students in the higher grade levels have had less time to have math skills reported into their records than have migrant students at mid-level grades.

The third factor is the grade level distribution of skills in the base math skills system. There are simply fewer math skills in the base system that apply to the higher grades as there are skills which apply to the lower and mid-level grades as was demonstrated in EXPERIMENTAL SET I.

The "current + previous" skills count contains a restricted cumulative component which consists of the previous skills. Because of this component, the same remarks concerning the cumulative skills may be applied to the current + cumulative skills.

b. Plots C3 and C4 are presented to illustrate one aspect of the relationship between entry age and grade level. Plot C3 shows the range of (90%) of entry ages associated with each grade level on the tests. Plot C4 shows the range of (90%) of the grade levels associated with each entry age on the tests. Notice that as grade level increases, the range of entry ages also increases (from Plot C3); whereas, as entry age increases, the range of grade levels decreases (from Plot C4). This effect is illustrated in Plot C5 as degrees of overlap of distributions of grade levels (see Plot C5-1) as entry age is varied from a value of 8 to 16 versus degrees of overlap of distributions of entry ages (see Plot C5-2) as grade level is varied from a value of 2 to 10.

Plots C5-1 and C5-2 seem to suggest that increased grade level (of test scores) might not add as much "grade level" effect to test performance degradation as increasing entry age (of test scores) might add "entry age" effect. Increasing from grade level 12 to grade level 10 added 56.75% of entry ages that were all greater than the maximum entry age in the grade level 2 distribution. Conversely, increasing from entry age 8 to entry age 16 added only 20.33% of grade levels that were higher than the maximum grade level in the entry age 8 distribution. This relative degree of effect would be true only if a "unit" of grade level had the same effect as a comparable unit of entry age. Plots C6 and C7 demonstrate the relative amounts of increase/non-increase in each variable as a function of increases in the other variable.

Plots C8 and C9 present the relationships being discussed in a more familiar manner. Notice in Plot C9 that:

*the means increase at the same rate,
*the standard deviation of entry ages increases at the same rate as does the mean entry age,
*the standard deviation of grade level decreases with increases in mean grade level.

The coefficients of variation in Plot C8 simply portray the relationship between the means and standard deviations more directly.

VARIABLES: MEAN NUMBER OF MATH TESTS PER STUDENT = GRADE LEVEL

The previous section presented evidence that math skills and math tests are independently distributed at each grade level. This is no longer a primary concern. However, there are two effects that varying numbers of test scores per student at different grade levels might have that are of concern to the present study.

Effect One: If students who are tested multiply on the same test at the same grade level, the increased testing activity might affect their test scores. The effect could be positive (developing test sophistication) or negative (developing test fatigue and frustration). Some educators have expressed concern about this effect mostly in its negative aspect.

"Regular" students are not ordinarily administered standardized tests each year. If the data shows that migrant students are tested at a greater frequency than are regular students, then comparisons between the standardized test scores of migrant students and those of "regular" students rest upon the assumption that differential testing frequency has no effect on test scores. This would be an assumption that MUST BE VALIDATED EMPIRICALLY FOR THE ENTIRE MIGRANT STUDENT POPULATION. This is but one more reason that such group comparisons were avoided by the present study.

Effect Two: Part II of the present study has established, unequivocally, that migrant students' performance on standardized (CAT and CTBS) math tests decreases with increases in grade level. If the frequency of math tests per student increases across grade level this will have the effect of loading "pooled" scores (scores averaged across grade levels) with those representing the higher grade levels and consequently with lower scores.

The proportional distribution of math tests across grade levels will be investigated with the two above described effects as the focus of concern.

(Type of analysis: ANOVA)

Table J3 shows that the mean number of math tests per student differs significantly across grade level for all populations.

The chart below shows, for each population the highest mean number of math tests reported in any grade level, the lowest mean number of math tests reported in any grade level and the percent of change in means.

NOTE: The grade level at which the highest/lowest mean occurred is in parentheses.

Present in each record making up the population	HIGH MEAN #TESTS	LOW MEAN #TESTS	PERCENT OF CHANGE
1 math skills	1.53 (5)	0.85 (11)	44.44
2 math skills or read skills	0.98 (2)	0.83 (11)	15.30
3 math skills and read skills	1.01 (3)	0.82 (12)	18.81
4 math test	1.07 (1)	1.01 (12)	5.60
5 math test or read test	0.99 (2)	0.83 (12)	16.16
6 math test and read test	1.07 (1)	1.01 (12)	5.60
7 math skill and math test	1.11 (12)	1.05 (11)	5.71
8 math skill or read skill and math test	1.09 (9)	1.04 (11)	4.58

The following observations are in order regarding Table J2 and the above chart of means.

(Overall)

i. The incidence of testing does appear to be greater than in the population of "regular" students. Population 4 represents the population that would be used for comparisons with other groups.

For population 4, the chart below presents, for each grade level (top row), the means (middle row) and standard deviations (bottom row).

1	2	3	4	5	6	7	8	9	10	11	12
1.07	1.07	1.06	1.06	1.06	1.05	1.05	1.04	1.05	1.04	1.02	1.01
0.27	0.26	0.26	0.25	0.23	0.23	0.23	0.21	0.22	0.30	0.15	0.12

(Within each population)

i. No particular pattern of magnitudes of differences within population high/low means may be observed.

ii. For all populations, the very large N's (large number of cases) within grade level produce a relatively small mean square error term which in turn produces a significant F ratio. This is enhanced by the fact that the degrees of freedom for the "model" is always low at 11 (i.e., total number of grade levels 1). As in the case of skills, this must not be taken to imply that the

differences among mean numbers of tests per student are not "real." Whether the MAGNITUDE of differences is large enough to be important is a question beyond the knowledge of the authors.

iii. The extremely low R-SQUARE values associated with these results indicate that the ANOVA model provides a very low accounting of the variability of number of tests per student within skill intervals.

iv. Even though each population showed significant differences, the R-SQUARE associated with the regressions suggests that the regression models provide a far better description of the data.

(Between populations)

Although there does appear to be clear differences between some of the population means, no particular pattern of differences was observed.

- -

(Type of analysis: REGRESSION)

i. For two populations, #4 (records having math test data) and #6 (records having math test data and reading test data), the curve of best fit is that of a negative linear regression. The best fit curve for the remaining populations are polynomial regressions. As the small magnitude of differences between the mean number of tests per student across grade levels would suggest, no dramatic results are to be expected. Neither the linear nor polynomial curves suggest any but the mildest change in the mean number of tests per student across the 12 grade levels. None of the curves would, by visual inspection alone, suggest non-zero regression terms. Plots J3 and J4 are included to illustrate the point.

ii. Population 4 (records with math tests) would represent the outcome of standardized testing in math. From the table presented previously, a change in mean number of tests per student from 1.07 to 1.01 (5.6%) tests across grade levels is the maximum change observed. Whether such a small change could affect test performance within that population is doubtful.

iii. Population 7 (records having math skills and math tests) is a population of particular interest to the present study. A change in means from 1.11 to 1.05 (5.71%) gives an absolute shift of 0.06 mean tests. The effect, if any, of such a change will be regarded as of no consequence to individual student test performance until evidence demonstrates otherwise.

iv. In conclusion, changes in the mean number of math tests per student as a function of grade level, though statistically significant, appear to be behaviorally insignificant in terms of affecting test scores WITHIN any of the eight populations under study. The authors would appreciate references to studies, data, anecdotes or any other form of information bearing on this topic.

- -

Summary of results for number of math tests = grade level

****
* From grade level 1-12, there is a very mild increase *
* in the number of math tests per student. *
****

VARIABLES:

MEAN NUMBER OF MATH TESTS PER STUDENT = MATH SKILLS/STUDENT

- -

This dependent/independent relationship variable is being examined as a precaution to insure that test scores are randomly distributed over students having differing amounts of skill data in their records.

To perform the analyses discussed below the math skills data in each record was cast in interval form as follows. All records having no math skills data were counted as interval zero. Twelve intervals of size three were created from 1 to 36. All records having a frequency of math skills greater than 36 were counted as "the interval whose midpoint is 38." All tables, plots and discussion use the midpoint of each interval as a reference (e.g., the interval of 7-9 is referred to as interval 8.) As is customary, all analyses performed on the interval used the midpoint as the value of the interval.

(Type of analysis: ANOVA)

Table J4 shows that the mean number of math tests per student differs significantly across skill intervals for all populations.

The following chart shows, for each population: the highest mean number of math tests reported in any skill interval, the lowest mean number of math tests reported in any skill interval and the percent of change in means.

NOTE: The skill interval at which the highest/lowest mean occurred is in parentheses.

Present in each record making up the population	HIGH MEAN #TESTS	LOW MEAN #TESTS	PERCENT OF CHANGE
1 math skills	1.18 (38)	0.96 (5)	18.64
2 math skills or read skills	1.18 (38)	0.81 (0)	31.35
3 math skills and read skills	1.12 (26)	0.95 (5)	15.17
4 math test	1.18 (38)	0.92 (0)	22.03
5 math test or read test			
6 math test and read test	1.21 (38)	1.03 (0)	14.87
7 math skill and math test			
8 math skill or read skill and math test			

The following observations are in order regarding Table J4 and the above chart of means.

i. From the preceding chart no particular pattern of magnitudes of differences among population high/low means may be observed.

ii. For all populations, the very large N's (large number of cases) within grade level produce a relatively small mean square error term which in turn produces a significant F ratio. This is enhanced by the fact that the degrees of freedom for the "model" are always low at 11 (i.e., total number of grade levels - 1). As in the case of skills, this must not be taken to imply that the differences among mean numbers of tests per student are not "real." Once again, the authors have no way of knowing whether the MAGNITUDE of differences is large enough to produce test performance on the part of individual students.

iii. The extremely low R-SQUARE values associated with these results indicate that the ANOVA model provides a very low accounting of the variability of number of tests per student within varying skill intervals.

iv. Even though each population showed significant differences, the R-SQUARES associated with the regressions suggest that the regression models provide a far better description of the data.

(type of analysis: Regression)

i. Table J4 shows that a linear regression model produces the best description of the data. There is, for each population, a positive relationship between the mean number of tests and the skill intervals represented by the records in the SKILLS/TEST data set.

ii. As the values of the slopes given in Table J4 indicated, any increase of the occurrence of math tests over math skill intervals is very slight. So slight that visual inspection of the plots fails to convey an impression of non-zero slope for the curves.

iii. Perhaps the most helpful conclusion to draw is that the mean number of math tests in a student's record DOES NOT DECREASE as the number of math skills in that student's record increases.

K. EXPERIMENTAL SET K

Migrant Student Performance on Standardized Math Tests as a Function of Varying Numbers of Math Skills in MSRTS Records. The Data Set Used for This Analysis was IDSHSK Which for Clarity Will be Called "the Skills/Test Data Set."

SKILLS/TEST data set description

The SKILLS/TEST data set represents all cases in which a student had math or reading test data reported under a grade level 1-12 enrollment. Table J1 in the ILLUSTRATIONS MANUAL gives essential parameters of the SKILLS/TEST data set and its derivation from the MSRTS data base. The reader is referred to the Appendix, Section C for details of the SKILLS/TESTS data set (identified as IDSHTDSK in the Appendix).

Three populations were drawn from the SKILLS/TEST data set. Each population was analyzed with respect to the percent of students scoring at or above grade level on standardized tests as a function of the number of math skills in the records.

1. Variables: (DEPENDENT VARIABLE = INDEPENDENT VARIABLE)

PERCENT AT OR ABOVE GRADE = MATH SKILLS INTERVAL
(see "5" Procedures below for a description of math skills intervals.)

The above relationship will be examined on current skills for each of the three populations listed below and on current + prior and cumulative skills for Populations 1. Unless otherwise specified, all references to skills or skill intervals mean "current" skills.

2. Populations:

1. Those records containing at least one math test.

2. Those records containing at least one math skill and at least one math test.

3. Those records containing at least one math skill or at least one reading skill and at least one math test.

3. Tests: For each population:

(using GLM simple regression model): Regression of percent at or above grade on skills interval.

(using GLM simple regression model): Regression of percent at or above grade on skills interval.

4. Hyp:

Hoa: For each population, the slope of simple regression of percent at or above grade on math skills interval = 0.

H1b: Slope not = 0.

Hob1: For each population, the linear term for the regression of the percent of scores at or above grade on math skills interval = 0.

Hob2: For each population, the quadratic term for the regression of the percent of scores at or above grade on math skills interval = 0.

H1b1: Linear term not = 0.

H1b2: Quadratic term not = 0.

5. Procedure:

To perform the analyses discussed below the math skills data in each record was cast in interval form as follows. All records having no math skills were counted as interval zero. Twelve intervals of size 3 were created from 1 to 36. All records having a frequency of math skills greater than 36 were counted as "the interval whose midpoint is 38." All tables, plots and discussion use the midpoint of each interval as a reference (e.g., The interval of 7-9 is referred to as interval 8.). As is customary, all analyses performed on the interval used the midpoint as the value of the interval.

6. Results:

See TABLES K1, K2, and K3 for summaries of the analyses performed on the three populations for current math skills. Tables K4 and K5 present the analyzed results for "current = prior" math skills and "cumulative" math skills over Population 1 (records having at least one math test but not necessarily a math skill).

The detailed tables of results for all analyses over all populations will not be presented because they are too numerous (12 grade levels X 2 types of analyses X 3 populations = 72 tables for current skills alone). These tables as well as all other data, tables and charts referenced directly or indirectly in this study have been catalogued and are on file for the authorized readers inspection at the MSRTS in Little Rock, Arkansas.

7. Discussion of results:

The results of each type of analysis will be reviewed next.

Results of LINEAR REGRESSION

i. For all three populations, all significant slopes in grade levels 1-8 inclusive are positive. For grades 9-12, only grade level 9 is best described by a simple linear regression model.

ii. Only the third grade scores fail to show a significant regression in all three populations.

iii. For all three populations, all significant slopes in grade levels 9-12 inclusive are negative.

iv. Plot K1 depicts the data of Population 1 as simple linear regressions. The abrupt change in the direction and magnitude of regression is dramatic in Plot K1.

v. Plot K3 depicts the data of Population 2 as simple linear regressions. The abrupt change in the direction and magnitude of regression observed for Population 1 (see Plot K1) is equally in evidence for Population 2.

vi. For Population 1, the linear regression results for "current + prior" math skills data (see Table K4) and for "cumulative" math skills data parallel the results for "current" math skills data for (see Table K1).

Results of POLYNOMIAL REGRESSION:

i. In none of the three populations does the quadratic term reach significance in grade levels 1-9 except for the eighth (8) grade in Population 2.

ii. In none of the three populations does the quadratic term reach significance in grade levels 9 or 11.

iii. In all three populations, the best description of the grade levels 10 and 12 is provided by a polynomial regression model.

iv. Plot K2 depicts the significant results of Population 1 as polynomial regressions.

v. Plot K4 depicts the significant results of Population 2 as polynomial regressions.

vi. For Population 1, the polynomial regression results for "current + prior" math skills data (see Table K4) and for "cumulative" math skills data parallel the results for "current" math skills data (see Table K1) except for grade level 9 which shows a significant quadratic term for other than "current" skills.

---

7. Discussion of Results:

To set the environment for the discussion for the results of EXPERIMENTAL SET K, it is necessary to review certain previous findings of this study.

Review of Previous Results

Results "a" From EXPERIMENTAL SET H

The percent of all students on whom math skills are reported decreases sharply in grade levels 9-12.

Results "b" From EXPERIMENTAL SET H

For those students on whom math skills are reported in grade levels 9-12 (this population 2 in the present analysis), the mean number of math skills reported per student does not decrease as grade level increases from grade level 1 to 12.

Results "c" From EXPERIMENTAL SET I

The skills being reported at grade levels 9-12 are inappropriate to those grade levels in the "normal" course of instruction, and reflect instruction aimed at teaching skills that students are intended to acquire at earlier grade levels.

From the above results, the following summary statement may be made (it has already been made in EXPERIMENTAL SET I).

****
* The skills data reported for grade level 9-12 migrant *
* students represents a very specific subset of (9-12) *
* students for whom instruction is conducted that is aimed *
* at teaching those math skills that are ordinarily acquired *
* at earlier grade levels. *
****

End of Review of Previous Results

Discussion of results continued.

(results within populations)

1. As Plots K1 and K3 show, there is a dichotomy of performance based on grade level within each population. In general, student performance increases across grade levels 1-8 and decreases across grade levels 9-12.

   ii. The increase in performance across grade levels 1-9 is a fairly uncomplicated linear function. This is not to say that a causal relationship between skills data in records and test performance is being asserted. There are many other factors which could take causal precedence over the amount of skills data in students records. For example, it may be that schools reporting the most skills are more conscientious and also conduct better compensatory programs. It may be that having skills data in students' records is not, per se, the dominant factor but that having a "structure" of universal skill statements enhances continuity of instruction from the students' perspective. At any rate, the incontestable finding is that migrant student performance on standardized math tests does rise as a statistical function of the number of math skills in student records.

   iii. Seen as a simple linear relationship, the decline in performance in grade levels 9-12 invites explanation. The results reviewed from previous experimental sets lead to the formation of the following multi-part hypothesis.

```
**************************************************************************
*  Hypothesis 1, Part 1:                                                 *
*  The decline of standardized tests scores among grade level 9-12       *
*  students is a reflection of the fact that those students are being    *
*  tested over skills for which none or little usable instruction        *
*  has been provided.                                                    *
*                                                                        *
*  Hypothesis 1, Part 2:                                                 *
*  If grade level 9-12 migrant students were to be tested over           *
*  material that was commensurate with the skills data in their          *
*  records, then the relationship between test performance and           *
*  number of skills in student records would be positive as it           *
*  is for students in grade levels 1-8.                                  *
**************************************************************************
```

The test scores used in this analysis indicate that many migrant students are placed in higher grade levels than is appropriate to their math skill level. This statement is not a criticism of the placement function. There are many factors which must be considered in the placement of migrant students that are equally as important as readiness for secondary math! The point remains that a student (migrant or not) sitting through an hour a day of high school math subjects WITHOUT POSSESSING THE PREREQUISITE SKILLS cannot be said to be

receiving USABLE instruction in that subject. To administer grade level 9-12 standardized math tests whose level is determined by the grade level of the migrant student is at best meaningless and at worst destructive of the desire for the student to "keep on trying."

The scores from tests administered to students who have received no usable instruction over the subject matter being tested are surely not a useful measure of achievement, for these students have had no opportunity to achieve. Why are grade 9-12 students not tested in such subjects as molecular biology, differential calculus and particle physics? They are not tested simply because they have received no USABLE instruction in such subjects, and nothing useful would be served by the massive numbers of low scores that would result from such tests. Such a testing program seems absurd. Yet this is exactly the testing environment for many migrant students and particularly for those shown in the data being analyzed.

At best these test scores give a rough idea of how much opportunity to learn has NOT been provided each migrant student by the total circumstances of his or her life. (i.e., For 9-12 migrant students, the scores identify those grade levels at which the student has NOT received enough usable instruction to warrant being tested.). Such an illogical process could be avoided by using a pre-test screening technique to assess the appropriate level at which to test the student.

iv. Remembering that the math skills being reported for grade level 9-12 students are those skills appropriate to grade levels 1-8, and remembering that the percent of students on whom skills are reported drops dramatically in grade levels 9-12, the best interpretation of the curves in Plots K1 and K3 is that they show:

the extent to which the 9-12 migrant students were unprepared for the level of test they were administered,
and
the extent to which they were unprepared to USE the math instruction conducted at the 9-12 grade level.

An eleventh grade migrant student who had in excess of 25 grade level 1-8 skills in whole numbers and fractions reported in his or her record (as an eleventh grader) was clearly less able to USE (at the beginning of the academic year) instruction in high school geometry or algebra than is the eleventh grader who needed no 1-8 math skills (great assumptions about differential skills reporting aside!). As will be seen in the following, the two interpretations stated above hold in principle but must be modified functionally.

vi. Although the data for grade levels 9-12 may be described by simple linear curves, a much better description is provided by a second order polynomial. This means that the relationship between the number of math skills reported per student and their performance on the math tests in grade levels 9-12 is a more complicated phenomena than it is in grade levels 1-8.

Plots K2 and K4 show a remarkable decline of scores up to a level of skills and then an upsurge of scores as the number of reported math skills increases! If the interpretation of the decline of scores given in paragraph vi preceding is correct, instruction aimed at providing the grade level 1-8 skills to the grade 9-12 students is successful! Up to a point, the number of skills reported indicates need; beyond that, the number of skills reported indicates need being met. These results lead to the following interpretive multi-part hypothesis.

```
**************************************************************************
*    Hypothesis 2, Part 1:                                               *
*    The students on whom skills are reported are generally those        *
*    students lacking the prerequisite skills for grade level 9-12       *
*    math subjects.                                                      *
*                                                                        *
*    Hypothesis 2, Part 2:                                               *
*    Except for grade level 11, the rise in test scores that occurs      *
*    beyond a certain number (17-29) of reported skills reflects the     *
*    success in providing the prerequisite skills.                       *
**************************************************************************
```

(Discussion of results between populations)

i. If it is true that grade level 1-8 test scores are positively related to the number of skills in student records, then scores should rise within populations as a function of the number of skills present for those grade levels. This has been shown true. In addition, scores should rise faster in populations having a greater proportion of grade level 1-8 students with math skills. From Table J1, we know that of the 131,868 records of population 1 65,134 (about half) contain no math skills data. However, from EXPERIMENTAL SET J we know that most of the records having math skills occurred in grade levels 1-8. Thus we know that more than half of the grade level 1-8 records that contained math test data also contained math skills data. Hence population 2 had a greater proportion of records containing math skills data than did population 1 and should have greater slopes.

This expectation is confirmed by comparing the slopes of Population 1 with those of Population 2. Tables K1 and K2 show that on a grade by grade comparison each grade level 1-8 slope for population 2 greater than for Population 1 (no tests of significance was run on these differences).

ii. The differences between the grade level 9-12 curves for Populations 1 and 2 are apparent from Plot K2 and K3. In grade levels 9-12 the math skills being reported for both populations are essentially grade level 1-8 skills. Therefore, no differential effect between populations was to be expected. The only change that was not predicted by the interpretive hypotheses proposed herein is the failure of grade level 9 to reach significance in Population 2. To offset this however, grade level 9 Population 2 scores also failed to reach significance as a negative linear function of skills. Because the scores for grade level 9, Population 2, were accounted by neither a linear not a second degree polynomial, it is shown in Plot K5.

TABLE OF CONTENTS

PART IV

EXPERIMENTAL SETS L-O

EXAMINATION OF THE READING SKILLS SYSTEM
AND
ITS RELATIONSHIP TO STUDENT PERFORMANCE

Introduction to the Skills Data.

An important note on calculating the mean number of skills per student.

		PAGE
L.	EXPERIMENTAL SET L Descriptive statistics of reading skills data in the SKILLS data set.	IV-1
M.	EXPERIMENTAL SET M Descriptive statistics of reading levels represented by the reading skills data in the SKILLS data set.	IV-5
N.	EXPERIMENTAL SET N Descriptive statistics of reading skills data and reading test data in IDSHTDSK.	IV-7
O.	EXPERIMENTAL SET O Migrant student performance and standardized reading tests as a function of varying numbers of reading skills in MSRTS records. The SKILLS/TEST data set was used for this analysis.	IV-21

Note 1: When inspecting the analysis of variance tables presented herein, always refer to the TYPE III values (as opposed to TYPE I).

Note 2: Most of the data plots represent very large samples relative to the population. Vertical differences between data points may often appear not to be significant when in fact they are.

Note 3: Simple regressions are normally run for A on B and B on A. For this reason, the independent/dependent variables involved in the regression analyses are simply specified as members of the set (A, B).

Note 4: All testing is done at alpha = .05.

****
* All tables and plots referenced herein will be found in the *
* accompanying ILLUSTRATIONS MANUAL. *
****

Introduction to the Skills Data

Part II preceding examined some of the overall relationships between the migrant program and student performance on standardized tests. Part IV will examine the reading skills system data and the relationship between that data and student performance on standardized tests. Detailed information regarding the formation of the data sets used in PART IV may be found in the APPENDIX, Section C, Populations Selected for Use in Studying the Effects of the Skills System on Standardized Scores in Math and Reading. A summary of the numbers of skills and/or tests making up the populations studied will be presented in each experimental set.

An Important Notice on Calculating the Mean Number of Skills Per Student

When calculating the mean number of skills (by grade level) for each student, there is a fundamental decision to be made that affects the resulting values greatly. That decision is whether to include cases in which students have no (0) skills reported at a given grade level. The inclusion of the numeric value zero (0) will of course have the effect of producing lower means because it raises the number of cases (the denominator) without changing the total skills counted (the numerator). Because of the non-uniform implementation schedule of the skills reporting system across the national program, the lead author decided that the inclusion of ALL cases of no skills would not be meaningful and would introduce uncertainty as to the representativeness of the data sets. The author did not want to mix data from programs lacking skills reporting with data from programs having implemented skills reporting. In addition, the author wished to avoid the possibility of unknowingly introducing underlying bimodal distributions of skills (i.e., frequent distributions of skills in which a non-trivial portion of the population has zero skills and the remainder has a more normal spread of skills). To exclude ALL cases if zero (0) skills reported would misrepresent the realities of the program in the opposite direction. A compromise solution consisting of the following rules was adopted:

> If math skills but no reading skills were reported for a student for a given grade level then a skills record was created and assigned the value of zero (0) for the number of reading skills reported for that student at that grade level.
>
> If neither reading nor math skills were reported for a student at a given grade level then no skill record was created at that grade level for that student. This component of the rule has the effect of eliminating a zero (0) skills count for each record meeting the condition.

The rationale for use of the above rules is simply to eliminate from analysis those programs that probably had no skills reporting system in place at the time.

To insure that the presence or absence of zero (0) skills records did not bias the results unknowingly, each analysis was run in pairs - once including zero skills records and once excluding zero skills records.

L. EXPERIMENTAL SET L

Descriptive Statistics of Reading Skills Data in the SKILLS Data Set.

SKILLS Data Set Description

The SKILLS data set represents all cases in which a student had at least one math or reading skill reported under a grade level 1-12 enrollment.

EXPERIMENTAL SET L will examine the reading skills present in three populations drawn from the SKILLS data set. These populations are specified under item 2 populations, below.

1. Variables:

   Distribution statistics BY GRADE LEVEL including mean number of skills per student for each of the below three populations.

2. Populations:

   1. Those records containing at least one reading skill.
   2. Those records containing at least one math skill or at least one reading skill or both.
   3. Those records containing at least one math skill and at least one reading skill.

3. Tests: For each population:

   (Using GLM analysis of variance): Test of difference among mean number of reading skills at each grade level.

   (Using GLM simple regression model): Regression of mean number of reading skills on grade level.

   (Using GLM polynomial regression model): Regression of mean number of reading skills on grade level.

4. Hyp:Hoa:

   Within each population, there is no significant difference among the mean numbers of reading skills per student at different grade levels.

   Hlb:

   Difference among means.

Hob:

For each population, the slope of simple regression of mean number of reading skills on grade level = 0.

H1b:

Slope not = 0.

Hoc1:

For each population, the quadratic term for the regression of the mean number of math skills on grade level = 0.

H1c1:

Linear term not = 0.

5. Procedure:

*Distribution statistics for each population were reported.
*Tests as specified were run.

6. Results

See Table L1 for a summary of the analyses performed on all three populations.

Population 1: (Each record contained at least one reading skill.) (See Tables L5, L6 and L7.)

Significant

Difference of mean number of skills among grade levels.

Slope of regression of mean number of skills on grade level.

Not significant

Linear and quadratic term of (polynomial) regression of mean number of reading skills on grade level.

Population 2: (Each record contained at least one math skill or at least one reading skill or both.) (See Tables L8, L9 and L10.)

Significant

Difference of mean number of skills among grade levels.

Slope of regression of mean number of skills on grade level.

Linear and quadratic term of (polynomial) regression of mean number of reading skills on grade level.

Not significant

All tests produced significant results.

Population 3: (Each record contained at least one math skill and at least one reading skill.) (See Tables H11, H12 and H13.)

Significant

Difference of mean number of skills among grade levels.

Slope of regression of mean number of skills on grade level.

Linear and quadratic term of (polynomial) regression of mean number of reading skills on grade level.

Not significant

All tests produced significant results.

7. Discussion of results:

The data of all three populations is similar, and the same regression model provided the best "fit" for each population. The discussion of results will not be directed at each individual population except where such treatment is relevant.

---

i. The low R-SQUARE value of the analysis of variance indicates that some other independent variable (e.g., entry age) may be a stronger determinant of variability.

ii. Although the linear regression provides a good description of the data of all three populations, the best description of the data is provided by the polynomial regression (see Tables L1 and L5 through L13).

iii. In essence, these results may be rephrased as follows:

****
* For all students in the three populations, the mean *
* number of reading skills reported per student decreases *
* slightly from grade 1 until grades 7-9 and then *
* increases through grade level 12. *
****

The population at large has 406,559 records (Table L3). This number represents the maximum number of records that may contain at least one reading skill. Table L2 shows that 372,764 records do indeed contain at least one reading skill records. Thus it may be stated that about 92% (372,764 divided by 406,559) of all possible SKILL records contain one or more reading skills. Taking this analysis a step further, it may be calculated that of the 59,478 first grade records, 93.4% (55,610 divided by 59,478) possess reading skills data.

By similar reasoning, the percent of total possible cases containing reading skills may be computed at each grade level. A summary of such percents appears below.

Percent of Total Possible Cases Containing Math Skills at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
93.4	92.9	92.2	92.3	92.1	91.8	89.7	89.0	83.4	84.9	87.6	89.6

The above percents may be viewed as the percent of students on whom reading skills are reported at each grade level. Note that the percent of students having skills reported follows roughly the same function as does the mean number of skills per student (Plots L1-L3).

The preceding calculations and the results being reviewed lead to the following conclusion.

****
* Changes in the mean number of reading skills reported per student *
* across grade levels cannot be accounted for by changes across *
* grade level in the percent of students whom reading skills are *
* reported. *
****

M. EXPERIMENTAL SET M

Descriptive Statistics of Reading Levels Represented by the Reading Skills Data in the SKILLS Data Set.

SKILLS Data Set Description

The SKILLS/DATA set represents all cases in which a student had at least one math or reading skill reported under a grade level 1-12 enrollment.

1. Variables:

   The frequency of reading skills occurring under each reading level by grade level.

2. Populations:

   Those records in the SKILLS data set containing at least one math skill.

3. Tests: For each population:

   (Using GLM simple regression model): Regression of reading skill level of reported reading skills on grade level.

   (Using GLM polynomial regression model): Regression of reading skill level of reported reading skills on grade level.

4. Hyp:Hoa: The slope of the simple regression of reading skill level on grade level = 0.

   H1a: Slope not = 0.

   Hob1: The linear term for the regression of reading skill level on grade level = 0.

   Hob2: The quadratic term for the regression of reading skill level on grade level = 0.

   H1b1: Linear term not = 0.

   H1b2: Quadratic term not = 0.

5. Procedure:

   A grade level by reading level frequency table of reading skills was prepared for the population. Tests were run as specified.

6. Results:

See Tables M1, M2, M3 and Plot M1 in the ILLUSTRATIONS MANUAL.

i. The best description of reading skill level as a function of grade level is provided by the linear regression model (see Table M2).

ii. Reading skill level increases directly with increases in grade level (see Table M2 and Plot M1).

7. Discussion of results:

Reading level is a subjective measure assigned by teachers. The significance of reading skill statements is not intended to be behaviorally specific enough to convey a "level." The function obtained by the preceding regression analysis and the overall test results of PART II of this report raise the possibility that teachers are "influenced" by the grade level at which they are teaching when making estimates of the grade level of a reading skill being taught. Other results found by this study suggest that the reading levels reported with the reading skills are not synchronous with standardized reading test scores of the migrant students. Several possibilities are raised by these discrepancies two of which are stated below:

The assignment of level is accurate, and reading skills are being presented at levels far beyond the achievement levels of the students.

The assignment of level is inaccurate, and reading skills are being presented at the appropriate achievement levels of the students.

Of these two possibilities, the everyday "face to face" feedback available to the classroom teacher as a consequence of instruction would seem, to this author, to disfavor the former possibility and favor the latter.

In any case, the uncertainty surrounding the "levels" assigned to reading skills makes tenuous the interpretation of relationships between reading skills data in student records and student performance on reading tests.

N. EXPERIMENTAL SET N

Descriptive Statistics of Reading Skills Data and Reading Test Data in IDSHTDSK. For Clarity, the Data Set IDSHTDSK will be called "the Skills/Test data set."

SKILLS/TEST Data Set Description

The SKILLS/TEST data set represents all cases in which a student had math or reading test data reported under a grade level 1-12 enrollment. Table N1 in the ILLUSTRATIONS MANUAL gives essential parameters of the SKILLS/TEST data set and its derivation from the MSRTS data base. The reader is referred to the Appendix, Section C for details of the SKILLS/TEST data set (identified as IDSHTDSK in the Appendix) and Appendix, Section C for a definition of "current + prior" and "cumulative" skills.

A total of eight populations were drawn from the SKILLS/TEST data set and analyzed with respect to both the mean numbers of reading skills per student and the mean number of reading tests per student. In a very important sense, this is the world of skills and tests the world looks at in standardized testing. These populations are not equally important to the task at hand and will be discussed accordingly.

1. Variables:

   (DEPENDENT VARIABLE = INDEPENDENT VARIABLE)
   MEAN NUMBER OF READING SKILLS PER STUDENT = GRADE LEVEL
   MEAN NUMBER OF READING TESTS PER STUDENT = GRADE LEVEL
   MEAN NUMBER OF READING TESTS PER STUDENT = READING SKILLS/STUDENT

   The above three relationships will be examined for each of the eight populations listed below. Additionally the relation, mean reading skills per student=grade level, will be examined over "current + prior" and "cumulative" skills for Populations 4 and 7.

2. Populations:

   1. Those records containing at least one reading skill.

   2. Those records containing at least one math skill or at least one reading skill or both.

   3. Those records containing at least one math skill and at least one reading skill.

4. Those records containing at least one reading test.

5. Those records containing at least one math test or at least one reading test or both.

6. Those records containing at least one math test and at least one reading test.

7. Those records containing at least one reading skill and at least one reading test.

8. Those records containing at least one math skill or at least one reading skill and at least one reading test.

3. Tests:

For each population:

(Using GLM analysis of variance): Test of difference among mean number of reading skills at each grade level.

(Using GLM analysis of variance): Test of difference among mean number of reading tests at each grade level.

(Using GLM analysis of variance): Test of difference among mean number of reading tests as a function of varying numbers of reading skills in student records.

(Using GLM simple regression model): Regression of mean number of reading skills on grade level.

(Using GLM simple regression model): Regression of mean number of reading tests on grade level.

(Using GLM simple regression model): Regression of mean number of reading tests on varying numbers of reading skills in student records.

(Using GLM polynomial regression model): Regression of mean number of reading skills on grade level.

(Using GLM polynomial regression model): Regression of mean number of reading tests on varying numbers of reading skills in student records.

4. Hyp:Hoa:

Within each population, there is no significant difference among the mean numbers of reading skills (or tests) per student at different grade levels (or numbers of skills).

H1b:

Difference among means.

Hob:

For each population, the slope of simple regression of mean number of reading skills (or tests) on grade level (or numbers of skills) = 0.

H1b:

Slope not = 0.

Hoc1:

For each population, the linear term for the regression of the mean number of reading skills (or tests) on grade level (or numbers of skills) = 0.

Hoc2:

For each population, the quadratic term for the regression of the mean number of reading skills (or tests) on grade level (or numbers of skills) = 0.

H1c1:

Linear term not = 0.

H1c2:

Quadratic term not = 0.

5. Procedure:

*Distribution statistics for each population were prepared.
*Tests as specified were run.

6. Results:

See TABLES N2, N3 and N4 for summaries of the analyses performed on all eight populations.

The detailed tables of results for all analyses over all populations will not be presented because they are too numerous (3 pair of dependent/independent variables X 3 types of analysis X 8 populations = 72 tables). These tables as well as all other data, tables and charts referenced directly or indirectly in this study have been catalogued and are on file for the authorized readers inspection at the MSRTS in Little Rock, Arkansas.

7. Discussion of Results:

VARIABLES: MEAN NUMBER OF READING SKILLS PER STUDENT=GRADE LEVEL

(Type of analysis: ANOVA)

Notice from Table N2 that all differences between mean numbers of skills per student across grade levels were significant. However, the actual differences varied enormously depending on the population.

The chart below shows, for each population:

the highest mean number of reading skills reported in any grade level, the lowest mean number of reading skills reported in any grade level, and the percent of change in means.

NOTE: The grade level at which the highest/lowest mean occurred is in parentheses.

Present in each record making up the population	HIGH MEAN	LOW MEAN	PERCENT OF CHANGE
1 reading skills	6.19(1)	3.87(11)	37.47
2 math skills or read skills	6.60(1)	3.36(10)	49.09
3 math skills and read skills	7.28(1)	3.31(11)	54.53
4 reading test	4.60(1)	0.42(11)	90.86
5 math test or read test	4.60(1)	0.40(12)	91.30
6 math test and read test	4.67(1)	0.40)12)	91.43
7 read skill and read test	6.90(1)	3.88(11)	43.76
8 math skill or read skill and read test	6.65(1)	3.44)10)	48.27

The following observations are in order regarding Table N2 and the above chart of means.

i. For all populations, the very large N's (large number of cases) within grade level produce a relatively small mean square error term which in turn produces a significant F ratio. This is enhanced by the fact that the degrees of freedom for the "model" is always low at 11 (e.g., total number of grade levels-1). This is not to imply that the differences among means are not "real."

ii. The extremely low R-SQUARE values associated with these results indicates that the ANOVA model provides a very low accounting of the variability within grade level.

iii. From the preceding chart it may be observed that the decrease in reading for Population 4 is much greater than the decrease in Populations 1 or 7. This difference is important because it implies that the population being tested in reading may not be representative of the population on whom reading skills are reported. This possibility will be investigated further.

(Type of analysis: REGRESSION)

i. All populations show a statistically significant negative regression of the mean number of reading skills reported per student on grade levels. This result is in keeping with the findings of EXPERIMENTAL SET L regarding the SKILLS data set-namely that the mean number of reading skills reported per student decreases across grade level.

ii. Reading skills data is distributed over the SKILLS data set and over the SKILLS/TEST data set in a similar fashion in that both populations show a decrease in the mean number of reading skills reported per student over grade level.

iii. Population 4 in TABLE N2 represents the skill records of migrant students who are tested in reading. Each record contains the results of one or more reading tests and may OR MAY NOT contain reading skills data. It has been shown (see Table N2 and Plot N1) that the mean number of reading skills per student in this population decreases across grade levels 1-12. This population represents the migrant student standardized reading test scores that normally would be used to represent the entire migrant population in any kind of comparison with other groups (e.g., regular students, students in other compensatory programs, etc).

Population 7 in Table N2 represents the skill records of migrant students who, at the same grade levels, have both reading test scores and reading skill data in their records. It has been shown (see Table N2 and Plot N1) that the mean number of reading skills per student in this population decreases across grade levels 1-12. Since any record in this population contain reading skills data, this collection of records is distributed in a similar fashion to Population 1 (see Table N2) and to the equivalent population, records containing reading skills, in the SKILLS data set. EXPERIMENTAL SET L has shown that the number of students on whom reading skills are reported does not decline from grade level 1 through grade level 12. A parallel analysis was performed on the SKILLS/TEST data set and is presented below.

iv. The SKILLS/TEST population at large (Population 5: records having reading skills data or reading test data) has 144,148

records. This number represents the maximum number of records that may contain reading skills data and reading test data. Appendix Figure E-4 in the ILLUSTRATIONS MANUAL shows that 76,515 records do indeed contain both reading skills and reading test data. Thus it may be stated that about 53% (76,515 divided by 144,148) OF ALL POSSIBLE SKILL/TEST records contain both reading skills and reading test data. By similar reasoning, the percent of total possible cases containing both reading skills and reading test data may be computed at each grade level. A summary of such percents appears below.

Percent of Total Possible Cases in the SKILLS/TEST Data Set Containing Both Reading Skills and Reading Test Data At Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
63.3	62.6	63.0	62.8	62.6	59.2	49.6	45.9	20.4	13.4	10.5	10.0

The above percents may be viewed as the percent of students on whom reading skills and reading tests are reported at each grade level. Note that the percent decreases sharply in grades 9-12. This percentage decrease could come from either of three sources:

a decrease (over grade levels) in the number of records containing reading skills data, or

a decrease (over grade levels) in the number of records containing reading test data, or

a change (over grade level) in the joint distribution of tests and skills.

The actual source of the percentage decrease (over grade level) in the number of records containing both reading skills and reading test data will be examined in paragraphs vii and viii following.

v. Based on the same rationale used in paragraph iv preceding, the percent of all possible cases in the SKILLS/TEST data set containing reading skills was computed at each grade level and appears below.

Percent of Total Possible Cases in the SKILLS/TEST Data Set Containing Reading Skills at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
66.6	64.7	64.3	63.9	63.6	60.1	50.6	46.8	20.8	13.7	10.6	10.1

It may be observed that the drop in the percent of cases containing reading skills roughly parallels the drop in the number of cases having reading skills and reading tests in the percent summary appearing in paragraph iv preceding.

vi. Based on the same rationale used in paragraph iv preceding, the percent of all possible cases in the SKILLS/TEST data set containing reading tests was computed at each grade level and appears below.

Percent of Total Possible Cases in the SKILLS/TEST Data Set Containing Reading Tests at Each Grade Level

1	2	3	4	5	6	7	8	9	10	11	12
94.4	95.8	97.0	97.3	97.2	97.3	97.0	96.9	96.0	95.8	97.0	97.4

It may be observed that the distribution of the percent of cases containing reading tests is not characteristic of the distribution of the number of cases having reading skills and reading tests in the percent summary appearing in paragraph iv preceding.

vii. If two variables are (statistically) independent, then their joint occurrence is given by the product of their individual occurrences. Based on this principle, the probability of joint occurrence (expressed as a percent) of reading skills and reading tests, as two independent variables, was calculated at each grade level and is presented below. Beneath each calculated estimate of joint occurrence is the percent that the estimate deviates from the actual data as presented in paragraph iv preceding.

Estimate of the Percent of Total Possible Cases in the SKILLS/TEST Data Set that would Contain both Reading Skills and Reading Tests at Each Grade Level if Reading Skills and Reading Tests were Distributed Independently

1	2	3	4	5	6	7	8	9	10	11	12
62.8	62.0	62.4	62.1	61.8	58.4	49.1	45.4	20.0	13.1	10.3	9.9
0.7	1.0	0.9	1.0	1.2	1.4	0.9	1.2	1.7	2.2	2.1	1.0

Discounting rounding error (all values were truncated) two conclusions seem to be in order:

One: Reading skills and reading tests are independently distributed over the SKILLS/TEST data set.

Two: If skills and tests are not independently distributed, then no pattern of non-independence seems to emerge. The greater percentages of discrepancy in the higher grades is probably related to sampling error induced by the smaller number of cases at those grade levels.

In effect, the above chart eliminated non-independence between the reading skills and reading tests as a source of the drop in the joint occurrence of reading skills and reading tests in grade levels.

viii. Summary of findings and conclusions regarding the reading skills and reading tests in the SKILLS/TEST data set:

****
* There is a sharp decrease in grade levels 9-12 in the number *
* of records (students) having reading skills data. *
* *
* The number of records (students) having reading test data *
* does not decrease across grade levels 1-12. *
* *
* Reading skills data and reading test data are independently *
* distributed over the SKILLS/TEST data set. *
* *
* There is a sharp decrease in grade levels 9-12 in the number *
* of records (students) having both reading skills and reading *
* test data. *
* *
* Conclusion: *
* *
* This drop is most likely accounted for by a corresponding *
* decrease in the number of grade level 1-12 records containing *
* reading skills. *
****

****
* The mean number of reading skills reported decreases over *
* grade levels 1-12 uniformly (showing no particular drop *
* in grade levels 9-12) for all populations of the SKILLS/TEST *
* data set. *
****

- -

All analyses of mean number of reading skills as a function of grade level up to this point have been directed at "current" skills; that is, the reading skills reported for the enrollments occurring at each grade level. Two additional measures of skills were derived and will be examined for Population 4 (all records having reading tests) and Population 7 (all records having both reading skill and reading tests). These derived measures are "current + prior" reading skills and "cumulative" reading skills. These skills are as their name suggests. For a detailed description of their derivation, see the Appendix, Section C, Step 1 under "the derivation of IDSHTDSK (skills and test data)."

The curve of best fit for "current + prior" reading skills and for "cumulative" reading skills over Populations 4 and 7 were polynomials. These curves are shown in Plot N2 which for comparison sake also included best fit curves for "current" skills. A discussion of the interpretation and implications of these curves follows.

There are certain properties of the distribution of cumulative reading skills that hold for both Populations 4 and 7. The bottom three curves of Plot N2 show the distribution of the mean number of reading skills per student across grade levels for Population 4 (records having reading tests but not necessarily reading skills).

The top three curves of Plot N2 shows the distribution of the mean number of reading skills per student across grade levels for Population 7 (records having both reading tests and reading skills).

Note that the two curves displaying the cumulative skills (one for each population) run counter to the common sense notion that a cumulative quantity cannot decrease. The decrease in the cumulative skills as grade level increases beyond the fifth grade is a result of two factors which make the analysis of migrant student data different from that of regular students.

The first factor is the recognition that the reading skills being discussed are a part of the Migrant Education Program rather than the regular school program. Thus, skills accrue to a students record only if that student is enrolled in a migrant education program. The second factor is that of entry age-the age of the migrant student when first enrolled in a migrant education program. Entry age increases significantly with grade level. This means that migrant students in the higher grade levels have been members of migrant education programs for a far shorter period than have migrant students in the mid-level grades. Thus migrant students in the higher grade levels have had less time to have reading skills reported in their records than have migrant students at mid-level grades.

The current + previous skill count contains a restricted cumulative component which consists of the previous skills. Because of this component, the same remarks concerning the cumulative skills may be applied to the current + cumulative skills.

VARIABLES: MEAN NUMBER OF READING TESTS PER STUDENT=GRADE LEVEL

The previous section presented evidence that reading skills and reading tests are independently distributed at each grade level. This is no longer a primary concern. However, there are two effects that varying numbers of test scores per student at different grade levels might have that are of concern to the present study.

Effect one: If students are tested multiple times on the same test at the same grade level, the increased testing activity might affect their test scores. The effect could be positive (developing test sophistication) or negative (developing test fatigue and frustration). Some educators have expressed concern about this effect mostly in its negative aspect.

"Regular" students are not ordinarily administered standardized tests each year. If the data shows that migrant students are tested at a greater frequency than are regular students, then comparisons between the standardized tests scores of migrant students and those of "regular" students rest upon the assumption that differential testing frequency has no effect on test scores. This would be an assumption that MUST BE VALIDATED EMPIRICALLY FOR THE ENTIRE MIGRANT STUDENT POPULATION. This is but one more reason that such group comparisons were avoided by the present study.

Effect two: Part II of the present has established, unequivocally, that migrant students' performance on standardized (CAT and CTBS) reading tests decreases with increases in grade level. If the frequency of reading tests per student increases across grade level then this will have the effect of loading "pooled" scores (scores averaged across grade levels) with those representing the higher grade levels and consequently with lower scores.

The proportional distribution of reading tests across grade level will be investigated with the two above described effects as the focus of concern.

(Type of analysis: ANOVA)

Table N3 shows that the mean number of reading tests per student differs significantly across grade level for all populations.

The chart below shows, for each population the highest mean number of reading tests reported in any grade level, the lowest mean number of reading tests reported in any grade level, and the percent of change in means.

NOTE: The grade level at which the highest/lowest mean occurred is in parentheses.

Present in each record making up the population	HIGH MEAN #TESTS	LOW MEAN #TESTS	PERCENT OF CHANGE
1 reading skills	1.55(5)	1.26(12)	18.70
2 math skills or read skills	1.52(5)	1.25(12)	17.76
3 math skills and read skills	1.56(6)	1.30(12)	16.66
4 reading test	1.56(5)	1.29(12)	16.66
5 math test or read test	1.52(5)	1.25(12)	17.76
6 math test and read test	1.59(5)	1.28(12)	19.49
7 reading skill and read test	1.57(5)	1.28(12)	18.47
8 math skill or read skill and reading test	1.51(5)	1.29(12)	14.56

The following observations are in order regarding Table N2 and the above chart of means.

(Overall)

1. The incidence of testing does appear to be greater than in the population of "regular" students. Population 4 represents the population that would be used for comparisons with other groups.

   For Population 4, the chart below presents, for each grade level (top row), the means (middle row) and standard deviations (bottom row).

1	2	3	4	5	6	7	8	9	10	11	12
1.46	1.53	1.55	1.55	1.56	1.52	1.53	1.54	1.42	1.41	1.40	1.29
0.65	0.63	0.64	0.65	0.64	0.65	0.63	0.62	0.57	0.55	0.54	0.47

(Within each population)

i. No particular pattern of magnitudes of differences within population high/low means may be observed.

ii. For all populations, the very large N's (large number of cases) within grade level produce a relatively small mean square error term which in turn produces a significant F ratio. This is enhanced by the fact that the degrees of freedom for the "model" is always low at 11 (i.e., total number of grade levels -1). As in the case of skills, this must not be taken to imply that the differences among mean numbers of tests per student

are not "real." Whether the MAGNITUDE of differences is large enough to be important is a question beyond the knowledge of the authors.

iii. The extremely low R-SQUARE values associated with these results indicates that the ANOVA model provides a very low accounting of the variability of number of tests per student within skill intervals.

iv. Even though each population showed significant differences, the R-SQUARES associated with the regressions suggest that the polynomial regression models provide a far better description of the data.

(Between populations)

No particular pattern of differences between population means was observed.

(Type of analysis: LINEAR REGRESSION)

i. No slope for a linear regression model showed significance.

(Type of analysis: POLYNOMIAL REGRESSION)

ii. Each population showed a significant positive linear and negative quadratic term. This combination of signs produces an "umbrella" shaped parabola (see Plot N3).

iii. The R-SQUARE values of the polynomials signify a good accounting of the data.

Summary of results for reading tests/student = grade level

```
**************************************************************************
*    Migrant students appear to be tested in reading at a                *
*    higher frequency than regular students because migrant               *
*    students are tested each year.                                      *
*                                                                        *
*    For migrant students tested in reading, the frequency               *
*    of testing per student rises from 1.46 tests at grade               *
*    level 1 to a peak of 1.56 at grade level 5 and then                 *
*    falls to 1.29 at grade level 12.                                    *
**************************************************************************
```

VARIABLES:

MEAN NUMBER OF MATH TESTS PER STUDENT = MATH SKILLS/STUDENT

This dependent/independent relationship variable is being examined as a precaution to insure that test scores are randomly distributed over students having differing amounts of skill data in their records.

To perform the analyses discussed below the reading skills data in each record was cast in interval form as follows. All records having no reading skills data were counted as interval zero. Twelve intervals of size 3 were created from 1 to 36. All records having a frequency of reading skills greater than 36 were counted as "the interval whose midpoint is 38." All tables, plots and discussion use the midpoint of each interval as a reference (e.g., the interval of 7-9 is referred to as interval 8). As is customary, all analyses performed on the interval used the midpoint as the value of the interval.

(Type of analysis: ANOVA)

Table N4 shows that the mean number of reading tests per student differs significantly across skill intervals for all populations.

The chart below shows, for each population: the highest mean number of reading tests reported in any skill interval, the lowest mean number of reading tests reported in any skill interval and the percent of change in means.

NOTE: The skill interval at which the highest/lowest mean occurred is in parentheses.

Present in each record making up the population	HIGH MEAN #TESTS	LOW MEAN #TESTS	PERCENT OF CHANGE
1 reading skills	1.58(5)	1.19(38)	24.68
2 math skills or read skills	1.18(38)	0.81(0)	31.35
3 math skills and read skills	1.12(26)	0.95(5)	15.17
4 reading test	1.18(38)	0.92(0)	22.03
5 math test or read test	1.58(5)	1.19(38)	24.68
6 math test and read test	1.21(38)	1.03(0)	14.87
7 reading skill and read test	1.61(5)	1.26(38)	21.73
8 math skill or read skill and reading test			

The following observations are in order regarding Table N4 and the above chart of means:

i. From the preceding chart no particular pattern of magnitudes of differences among population high/low means may be observed.

ii. For all populations, the very large N's (large number of cases) within grade level produce a relatively small mean square error term which in turn produces a significant F ratio. This is enhanced by the fact that the degrees of freedom for the "model" is always low at 11 (i.e., total number of grade levels - 1). As in the case of skills, this must not be taken to imply that the differences among mean numbers of tests per student are not "real." Once again, the authors have no way of knowing whether the MAGNITUDE of differences is large enough to produce test performance on the part of individual students.

iii. The extremely low R-SQUARE values associated with these results indicates that the ANOVA model provides a very low accounting of the variability of number of tests per student within varying skill intervals.

iv. Even though each population showed significant differences, the R-SQUARES associated with the regressions suggest that the regression models provide a better description of the data.

(Type of analysis: LINEAR REGRESSION)

i. Table N4 shows no significant slope for the linear regression model. The linear regression model does not provide a very good fit as may be seen from the R-SQUARE values.

(Type of analysis: POLYNOMIAL REGRESSION)

i. Each population showed a significant positive linear and negative quadratic term. This combination of signs produces an "umbrella" shaped parabola (see Plot N4).

ii. As the values of Table N4 indicate, changes in the occurrence of reading tests over reading skill intervals, when present at all, is very slight.

Summary of results for reading tests/student = grade level

```
**************************************************************************
*    No systematic variation in the mean number of                       *
*    reading tests per student over reading skills                       *
*    intervals is observed.                                              *
**************************************************************************
```

O. EXPERIMENTAL SET O

Migrant Student Performance on Standardized Reading Tests as a Function of Varying Numbers of Reading Skills in MSRTS Records. The Data Set Used for This Analysis was IDSHSK Which for Clarity Will be Called "the Skills/Test Data Set."

SKILLS/TEST Data Set description

The SKILLS/TEST data set represents all cases in which a student had math or reading test data reported under a grade level 1-12 enrollment. Table J1 in the ILLUSTRATIONS MANUAL gives essential parameters of the SKILLS/TEST data set and its derivation from the MSRTS data base. The reader is referred to the Appendix, Section C for details of the SKILLS/TEST data set (identified as IDSHTDSK in the Appendix).

A total of 3 populations were drawn from the SKILLS/TEST data set. Each population was analyzed with respect to the percent of students scoring at or above grade level on standardized tests as a function of the number of reading skills in the records.

1. Variables:

   (DEPENDENT VARIABLE = INDEPENDENT VARIABLE)
   PERCENT AT OR ABOVE GRADE = READING SKILLS INTERVAL
   (See 5. Procedures below for a description of reading skills intervals).

   The above relationship will be examined on current skills for each of the three populations listed below and on current + prior and cumulative skills for Population 1. Unless otherwise specified, all references to skills or skill intervals mean "current" skills.

2. Populations:

   1. Those records containing at least one reading test.

   2. Those records containing at least one reading skill and at least one reading test.

   3. Those records containing at least one math skill or at least one reading skill and at least one reading test.

3. Tests:

For each population:

(Using GLM simple regression model): Regression of percent at or above grade on skills interval.

(Using GLM polynomial regression model): Regression of percent at or above grade on skills interval.

4. Hyp:hoa:

For each population, the slope of simple regression of percent at or above grade on reading skills interval = 0.

Hlb:

Slope not = 0.

Hob1:

For each population, the linear term for the regression of the percent of scores at or above grade on reading skills interval = 0.

Hob2:

For each population, the quadratic term for the regression of the percent of scores at or above grade on reading skills interval = 0.

Hlb1:

Linear term not = 0.

Hlb2:

Quadratic term not = 0.

5. Procedure:

To perform the analyses discussed below the reading skills data in each record was cast in interval form as follows. All records having no reading skills data were counted as interval zero. Twelve intervals of size 3 were created from 1 to 36. All records having a frequency of reading skills greater than 36 were counted as "the interval whose midpoint is 38." All tables, plots and discussion use the midpoint of each interval as a reference (e.g., the interval of 7-9 is referred to as interval 8). As is customary, all analyses performed on the interval used the midpoint as the value of the interval.

6. Results:

See Tables 01, 02 and 03 for summaries of the analyses performed on the three populations for current reading skills. Tables 04 and 05 present the analysis results for "current = prior" reading skills and cumulative" reading skills over population 1 (records having at least one reading test but not necessarily a reading skill).

The detailed tables of results for all analyses over all populations will not be presented because they are too numerous (12 grade levels X 2 types of analysis X 3 populations = 72 tables for current skills alone). These tables as well as all other data, tables and charts referenced directly or indirectly in this study have been catalogued and are on file for the authorized readers inspection at the MSRTS in Little Rock, Arkansas.

The curve best describing each grade level having a significant regression term is shown in Plot 01 for Population 1 and in Plot 02 for Population 2.

7. Discussion of results:

The curves of Plot 01 suggest a "soft" (does not cover all cases) multi part hypothesis as follows:

****

Hypothesis 1, Part A.
Migrant students in regular classrooms who perform approximately at the same level as their regular classroom peers in reading and language receive little special instruction by program specialists. These students have relatively fewer skills reported in their records as do migrant students who perform less well in reading and language.

Hypothesis 1, Part B.
Migrant students in regular classrooms who perform below their regular classroom peers in reading and language receive special instruction by program specialists. These students have relatively more reading skills reported in their records (a natural by product of special instruction and attention) as do migrant students who perform less well in reading and language.

Hypothesis 1, Part C.
Up to some number of skills (established empirically at about 15-20 in Plots 01 and 02), the number of skills in a migrant student record is a direct indicator of the special instruction needed and therefore is also an (inverse) predictor of performance on standardized reading tests.

****

```
*         Hypothesis 1, Part D.                                          *
*         As the special instruction takes effect, the net result is     *
*         an increase or leveling off of standardized test scores.       *
**************************************************************************
```

As named, the above hypothesis is "soft" because it fails to account for all grade levels in Plot 01. Actually, the hypothesis fails to account only for grade levels 5 and 6 as the following discussion reveals. Grade levels 1 and 2 students perform relatively high on the reading tests as Plot E2 in the ILLUSTRATIONS MANUAL has shown in Part II of this report. Prior to grade level 3, there is not much opportunity for differences in educational experiences among MIGRANT students to show up as differences in test scores. This lack of discriminating power within the migrant student population makes the tests unable to detect effects of the skills system even if present in the first and second grade.

Further, for grade levels 1 and 2, at least 93% of the cases show 12 or fewer skills in student records (see Table 06). Within the skill intervals represented by this 93%, the test performance is fairly uniform because (causation?) all students are receiving attention in reading-not just those that are performing below their peers. Therefore increasing numbers of skills reported per student is not a signification of particular individual need as it is in grade levels 3 and beyond. The remaining 7% of cases (cases having higher skill intervals) do not provide enough "leverage" to influence a regression curve which is well anchored at the upper performance level.

Beyond grade level 9 there seems to be a very small proportion of students receiving special instruction in reading, if the reporting of reading skills is any measure of this activity at those grade levels (see Table 06). There is simply not enough of a "spread" of numbers of skills per student at the higher grade levels to detect either a "need" for special instruction or the results of special instruction. One must consider that a migrant student requiring extensive assistance in reading and language at those grade levels will be unlikely to demonstrate improvement as a consequence of reading skills reporting.

Those directly involved in the teaching of reading and language would provide a far better explanation of these findings than the authors. A consensus of expert judgement would assist in understanding the dynamic interplay among program, test performance and skills information in reading. The authors realize that many different ideas (some contradictory-some not) would probably result from a call for expert opinion. However, such ideas would represent a valuable guidepost in developing techniques for getting relevant empirical data on the subject for the purpose of quantification as well as verification.

APPENDIX TO TEXT

APPENDIX

Table of Contents

		Appendix Page
A.	Populations and Samples	1
	1. Data that is and is Not Reported to the MSRTS	1
	2. Populations Represented by the MSRTS Data Base	4
B.	Populations Selected for Use in Studying the General Effects of the Migrant Education Program on Standardized Scores in Math and Reading	8
	1. Identification of Populations	8
	2. Cardinality of Populations	9
C.	Populations Selected for Use in Studying the Effects of the Skills System on Standardized Scores in Math and Reading	10
	1. Identification of Populations	10
	2. Cardinality of Populations	10
	3. Procedures for Deriving Data Sets See page 13 for definitions of "current + prior" skills and "cumulative" skills	11
D.	SAS Analysis Module References	15
	1. SAS Module Reference for Tables and Plots	15
	2. SAS Modules used for the Analyses of Skills Data	20
E.	Glossary of Terms	29

A. Populations and Samples

1. Data that is and is Not Reported to the MSRTS

The design of the MSRTS intends that certain events in the lives of migrant students trigger the transmission of data describing attributes of those events to the MSRTS if the student is under the umbrella of a migrant education program component.

Example of event	Example of resulting data
student enrolls in a school	enrollment data
student takes a standardized test	test data
student acquires a reading skill	reading skills data
student receives a polio immunization	immunization data

Several aspects of this situation help elucidate the boundaries of the present study and are worth noting.

Note 1

The events triggering the transmission of data to the MSRTS are ONLY those directly involved with the administration, instruction and health care of the student,

and

the data transmitted pertains only to the literal description of the administrative, instructional or health attributes of the events.

For example: The data used to describe the enrollment consists principally of a date, a status (an administrative matter) and a grade level (administrative and instructional). Whether the student is traveling with parents, appears confident, seems well nourished, is reasonably clothed against the elements, etc., are attributes of the enrollment that are not reported.

Another example: The test data reported does not include the number of times the student has already taken the same identical test (same version, level, etc.) during the past 12 months. The reported data also does not say whether the student gave up in frustration when barely into the test nor whether the student indicated an understanding of the test instructions.

Another example: The skills data reported does not indicate the amount of difficulty experienced by the student in acquiring the skill nor the

criterion used to decide that the skill had indeed been acquired. The skills data does not indicate whether a rise in confidence or self expectation occurred as a consequence of the skills mastery.

Note 1 is not an argument to include these attributes of student's experiences into the MSRTS data base but rather to recognize that there are many areas of student life, maturation and achievement that may be profoundly affected by the migrant education program that will never be reflected in the MSRTS data base simply because such is not the designed intent. It follows that inferences concerning the migrant education program based solely on a study of data in the MSRTS data base must be confined to those attributes of student life that are DESIGNED to be represented in the data base.

Note 2.

In addition to that data not designed into the MSRTS data base out of choice, there is data that is not designed into the data base simply because of the impossibility (and/or great cost) of collecting it. For example, when a migrant student leaves a school the student (or parents) usually do not initiate a formal withdrawal procedure. In such cases, the school may finally surmise the fact of withdrawal but has no way of knowing the reasons the student left school. The length of time that the student is out of school before enrolling in another school can only be deduced from the MSRTS data base based upon a subsequent enrollment in a school having a migrant program! Such deductions are accurate only if the subsequent enrollment is the first and only school the student attended since leaving the last school having a migrant program. If no subsequent enrollment in a migrant program school is received, there is no way of knowing the status of the student (e.g., "settled out" in a non-migrant program school, dropped out, ill, deceased, moved out of the United States, joined the armed forces, etc.). In short, the MSRTS may tell us many things that do happen to migrant students but, unlike the case with regular students, very little may be inferred from the ABSENCE of information concerning migrant students.

Note 3.

Not all schools having migrant students have a migrant education program. Data describing the administrative, instructional and health events of migrant students passing through these schools will not become a part of the MSRTS data base. Even in schools operating migrant education programs, not all data designed to be reported gets reported.

The foregoing material has discussed data that is and is not reported to the MSRTS. In a strict sense, this data may be described as "samples" in the following manner:

EVENT SAMPLES

Of the set of all the events that transpire in the life of migrant students, a non-random sample of events in that set triggers data being

transmitted to the MSRTS. The purpose of the transmitted data is to describe selected attributes of the events.

ATTRIBUTE SAMPLES

Of the set of all attributes of an event that could be reported to the MSRTS only a non-random sample of such attributes is designed to be reported to the MSRTS.

DESIGNED ATTRIBUTE SAMPLES

Of the set of all attributes designed to be reported to the MSRTS a (large) non-random sample is reported.

It may be stated that a sample (of designed attributes) of a sample (attributes) of a sample (of events in the lives of migrant students) is reported to the MSRTS. Overtime, the allowable accumulation of these samples makes up the MSRTS data base. It may also be stated that such a convoluted set of relationships and terms will lead to non-communication if applied throughout the present study.

In a sense, the term "population" may be substituted for the term "set" in the descriptions of the above three samples. This substitution might, however, suggest an inappropriate relation between the POPULATION AND THE SAMPLES. In a statistical sense, the terms "population" and "sample" are used in the following manner:

```
**************************************************************
*  A population is an aggregate about which we wish to know  *
*  certain properties but because of theoretical and/or      *
*  practical reasons, we must infer said properties from a   *
*  sample representing the population.                       *
**************************************************************
```

This definition would not permit us to say, in a statistical sense, that the sample of events selected to trigger the transmission of data to the MSRTS will yield information concerning the population of ALL events in the lives of migrant students. (The designers of the MSRTS had no such purpose in mind anyway.) Accordingly, we cannot say that the collection of attributes selected to report an event can be used as a sample to yield information about ALL attributes of such an event. Given that the MSRTS data base does not represent a sample in the sense just described, we must resolve two questions before we may proceed with a useful analysis of data.

Question 1: What population does the MSRTS data base represent?
Question 2: In what sense is the MSRTS a "sample" of this population?

Answers to these questions are the subject of the following section.

2. The Populations Represented by the MSRTS Data Base

There are roughly two approaches to performing a study of data in the MSRTS data base although the line between the two is not hard and fast. The first is a "time slice" study which takes all data at a point in time. The second is a "longitudinal" study which may compare successive "time slices" in some controlled manner. The present study chose the "time slice" approach. (The reader must bear in mind that since the MSRTS data base contains historical data, the choice of a "time slice" approach does not rule out the study of individuals or groups over time!) The term "time slice" refers to the MSRTS data base and not to the individuals represented in the data base. As of a certain date, the entire MSRTS data base was read and "reduced" into data sets containing data of interest to the study. Although the data sets might be used repeatedly to generate additional groupings of data, the MSRTS data base was read once and only once after final decisions had been made concerning the design and content of the required reduced data sets. Thus, the "time slice" in fact represents a "snapshot" of the data base at a particular point in time. This snapshot exists exactly once in the lifetime of the MSRTS. This precise snapshot never existed before and will never exist again. We may now state a meaning of the terms population and sample that satisfy the intent of statistical concepts and will be of use to the present study. For simplicity, it is tempting to say that all possible snapshots of the MSRTS that can ever exist are the population of which our snapshot is a sample. Unfortunately, this statement still obfuscates the situation.

This study wishes to confine inferences to other possible time slices to a smaller time zone because the essential operating character of the MSRTS has changed since its inception. For example, the data item, grade level, was not a part of the data base for several years. Tahus any analysis of data requiring grade level to be present must be confined to those snapshots which could have included grade level. Characteristics of the migrant education program itself have changed over the years. For example, the introduction of the skills system may have changed the relationships between test scores and other factors. These considerations lead to specific restrictions in the use of the term population as stated below.

```
**********************************************************************
*  The population being studied shall consist of all snapshots       *
*  that occur within a time zone in which both the MSRTS and         *
*  migrant education are similar to their operating                  *
*  characteristics when the sample snapshot used in the present      *
*  study was made (early 1986).                                      *
**********************************************************************
```

The above is not a very strict definition because it leaves the terms "operating characteristics" and "similar" undefined. For practical purposes, the past boundary of the time zone is about early 1982 and the reader is responsible for the future. Informally, it may be stated that

inferences drawn from the sample snapshot used in the present study may be applied to the population of data represented by the MSRTS from early 1982 until such time as the migrant education program or MSRTS is changed in ways to invalidate such inferences. The reader should be cautioned of a hidden assumption in this assertion. The snapshot used in the present study was not chosen at random. Therefore, the validity of the preceding assertion rests upon the assumption that the snapshot taken will represent the population as well as any other snapshot. At first this assumption may be somewhat disquieting. However, two observations should reassure the reader. First, the inertia of the population is substantial. It simply does not change its essential character unless there are changes in the operating characteristics of the program or the MSRTS. Second, the sample itself is large, not in regards to all possible samples, but in its complexity and data content. Despite the large volume of data in the sample, it was never necessary to "sample from the sample" to perform the statistical analyses. That is, all data in the snapshot that pertained to each hypothesis was used in the analysis of the hypothesis.

Whereas a satisfactory technical meaning of the terms "sample" and "population" may have been devised, they may still lead to erroneous assumptions. Traditional use of the term "sample" leads one to think of drawing a lesser collection of data from a larger population. As has been shown this is true only in the narrow and technical sense of having drawn one snapshot from a population of many. In reality, only one sample (the snapshot) was drawn for the entire study. All analyses performed during this study investigated programmatic effects on entire populations of data. For example, when the study refers to the reading scores of seventh graders, it is referring to ALL reading scores for ALL students on the data base that were in grade level seven when the test data was administered as of the time the snapshot was taken.

****
* To insure that the reader understands that ALL data in the *
* data base (at the time the snapshot was taken) that is *
* relevant to each analysis is used, the term sample will not *
* be used to report the results herein. *
****

The term "population" will be used in place of sample for the reasons discussed below.

Although in a strictly technical sense the population under study is a collection, bound by time, of snapshots of the entire data base, such jargon conveys little to the reader about each particular analysis undertaken. The snapshot is, after all, a collection of data which may tell us a little about how students are getting along scholastically and how various aspects of the migrant education program may be affecting their performance!

This is the purpose of the entire study. Rigor is of course a must when using statistics and an understanding of the formal properties of the population and samples must be stated. These objectives have been accomplished in the foregoing. The task at hand is to fashion language tools that suggest the proper "flavor" to the reader and do not invite erroneous interpretations despite being technically correct. The snapshot used in the analysis includes all data on all migrant students who had received at least one program service prior to the time the snapshot was taken. This is a view of the entire "student population" in the snapshot. But, what if an analysis must examine some property of "all students in the snapshot who had at least one reading skill in their records?" What shall this collection of data be said to represent? We have already ruled out the use of the term "sample" to be applied to such a collection. Technically it could be called a "sub-population" because it is lesser (usually) than the total number of students in the snapshot population. This term, however, will be avoided because it is needlessly clumsy (as is "target population," etc.) and it may be misleading. If indeed some kind of analysis of the records of all students having at least one reading skill is to be undertaken, then in the best sense of the term. That collection of students (data, records, etc.) shall be called:

****
"The Population"
****

The reader will encounter the term population frequently in this report because understanding the effects of the migrant education program on student performance begins with an understanding of the population of students that the program acts upon. For example, in studying how the skills system may affect student scores on certain tests, care must be taken to discover whether skills are reported on the population at large in the same proportions as skills are reported on students who are tested. There are many such inter-population factors to "control" during this study and so there are many populations to identify and study.

The manner in which the key terms discussed in the preceding will be used throughout this study are given below:

Data set	A collection of data that has some defined meaning in data processing terms. Synonymous with the concept of a file.
Snapshot	All data in the MSRTS data base, as of a given moment, collected in a file or data set.
Universe	All time slice snapshots of the MSRTS data base between early 1982 and such time as the essential

	operating characteristics of the migrant education program or the MSRTS are no longer the same as they were when the time slice snapshot underlying the present study was taken.
Population	All data in a snapshot (or in a data set derived there from) selected on the basis of having certain specified attributes. Thus many populations may be selected from a single data set.

B. Populations Selected for Use in Studying the General Effects of the Migrant Education Program on Standardized Scores in Math and Reading

From the MSRTS data base a data set was created using the following selection criteria. To be selected for inclusion in the data set, a student record must contain a legal value for:

* CTBS or CAT test data including:
  - year of administration 1983-1985
  - test score type of G.E. or percentile
  - test type of Arithmetic, Reading, or Oral Language
  - test score must be legal for score type
* status
* grade level associated with enrollment
* birth date.

The 351,011 records which were selected using the above criteria make up the General Effects data set.

1. Identification of Populations

From the General Effects data set, populations required for the various analyses were selected. The populations generated by the values of the major general effect variables are listed below:

Variable	Populations selected from the General Effects data set
Entry Age	Based on enrollment data and birth date, an entry age was calculated for each record.
Status	Status 4 was converted to Status 1, Status 5 to Status 2 and Status 6 to Status 3. These conversions resulted in three populations based on status.
Test Type	Five populations were generated on the basis of test type. * CTBS Arithmetic * CTBS Reading * CTBS Oral Language * CAT Arithmetic * CAT Reading
Grade Level	Twelve populations of scores were generated on the basis of grade level.
Entry Age = 6	All records having a (calculated) entry age of six years.

Minimal Exposure — All records whose test administration date occurred not more than six months following their initial enrollment in a migrant education program.

2. Cardinality of Populations

Population	Number of records
Entry Age CGMN2 15:03 5 MAY 86	1= 3,091 2= 5,920 3=10,265 4=19,105 5=29,403 6=72,114 7=49,372 8=35,287 9=29,608 10=23,839 11=20,567 12=15,856 13=11,997 14= 9,151 15= 6,310 16= 4,497 17= 2,612 18= 1,347 19= 499 20= 144 21= 20 22= 7
Status TDFR2 8:59 20 JUNE 86	1 = 66,714 2 = 59,627 3 = 224,670
Test Type TDFR2 8:59 20 JUNE 86	CTBS Arithmetic = 81,656 CTBS Reading = 85,084 CTBS Oral Language = 65,094 CAT Arithmetic = 58,024 CAT Reading = 61,153
Grade Level CGMN1 11:35 5 MAY 86	1=34,337 2=41,886 3=42,978 4=43,891 5=42,317 6=37,674 7=34,005 8=27,198 9=18,663 10=14,295 11= 9,666 12= 4,101
Entry Age = 6 Minimal Exposure EXFR1 9:20 6 AUG 86	49,372 12,727

C. Populations Selected for Use in Studying the Effects of the Skill System on Standardized Scores in Math and Reading

1. Identification of Populations

From the MSRTS data base four major data sets were formed from which to select populations for the analysis of the effects of the skills system on student scores.

Data set name	Data set content (By student by grade level)
IDSHTD	Test data
IDSHSK	All skills data on the data base
SKILLS	All grade level 1-12 skills from IDSHSK
IDSHTDSK	All test data from IDSHTD combined with all grade levels 1-12 skills for which test data existed in IDSHSK for the same student at the same grade level.

Figure Appendix C-1 (in the ILLUSTRATIONS MANUAL) describes the selection of data from the data base and the contents of these four data sets.

From two of these data sets (SKILLS and IDSHTDSK), various populations were selected for study individually and against other populations. The populations selected from the SKILLS data set are shown in Figure Appendix C-2 and those selected from IDSHTDSK are shown in Figure Appendix C-3.

2. Cardinality of Populations

The number of records making up each of the populations derived from the data sets involved in the analysis of skills (SKILLS and IDSHTDSK) are shown in Figure C-4. The number of skills and/or tests represented by the records in each of these data sets is the subject of Appendix Figures C-5 (for the SKILLS data set) and C-6 (for the IDSHTDSK data set). See the ILLUSTRATIONS MANUAL for Figures C-4, C-5 and C-6.

Aspects of these figures will be discussed extensively as the results of the skills analyses that are covered in PART III. Notice, however, that these populations are not mutually exclusive. The percents of each data set that the constituent populations do not sum to one hundred.

It will be helpful during the discussion of the results to make reference to populations of these data sets that are mutually exclusive. Such

populations are characterized in Appendix Figures C-7 and C-8. Notice in Figure C-7 that the value of EACH of the four variables (math skills, reading skills, math tests, reading tests) are FIXED by the selection criteria for each population. In Figures C-1 through C-5 only the variable(s) of interest are fixed in the selection criteria.

3. Procedures for Deriving Data Sets

The derivation of IDSHSK (skills data)

Step 1: At the time the "snapshot" of data used in this study was taken, the distribution of skills by type (regardless of associated grade level) in the MSRTS Data Base was as follows:

Math	4,529,734
Reading	2,846,569
Oral Language	1,042,253
Early Childhood	412,253
TOTAL	8,830,809

A subset of 8,418,556 skills was selected from the data base by excluding the early childhood skills. In constructing this subset, skills were grouped into one of two types as follows. All math skills retained the classification "math" regardless of their sub-classification within topic (e.g., Skills in the topic of geometry and skills in the topic of computation were both classified as "math" skills.). Reading and oral language skills were classified as "reading" skills. Since the dependent variable, being at or above on CTBS and CAT tests, is a measure lacking great power in discriminating among program nuances, it was decided to increase robustness by combining reading and oral language skills into a single category. The derived subset consisted of one record per skill. Each of the 8,418,556 records in this subset consisted of a student number, a skill description and the grade level the student was in during the enrollment under which the skill was reported.

Before the year 1983, grade level was not a required data item in the migrant student record, and most skills reported prior to 1983 had no grade level associated with them. A grade level classification of "missing" was assigned such records ("Missing" is not to be confused with the classification "ungraded" which is also present.). A total of 8,418,556 skill records were created in this subset. The following table summarizes their distribution:

SKILL TYPE

Grade Level	Math	Reading	Language	Total
Missing	1,782,281	1,249,051	312,837	3,344,169
Preschool	277,031	170,851	100,631	548,513 and K
Ungraded	28,991	23,357	11,962	64,310
1-12	2,441,431	1,403,310	616,823	4,461,564
TOTALS	4,529,734	2,846,569	1,042,253	8,418,556

SASSVFRQ 17:30 4/30 and SASSVFRQ 16:04 4/6 contain detailed skills distribution data.

Step 2: The records created in Step 1 above reduced into IDSHSK by collapsing all skills (by skill type) for a given student into a frequency count of each skill type (i.e., a frequency for math and a frequency for reading) by grade level. Thus, in IDSKSK, each student would have a record for each grade level in which at least one skill was reported. There were 656,067 "collapsed" skill records produced.

The record content of IDSHSK is illustrated in Appendix Figure E-9 of the ILLUSTRATIONS MANUAL.

The Derivation of the SKILLS Data Set

Of the 8,418,556 skills in IDSHSK, only 4,461,564 grade level 1-12 skills for math, reading and oral language went into the making of the SKILLS data set because before creating the SKILLS data set, the over-riding effect of grade level on test scores had been already established (in Experimental Sets A-F). The influence of grade level on test scores requires that analysis of skills be conducted by grade. For this reason, the ungraded and missing grade skills were excluded. Preschool and early childhood skills were excluded for two reasons. The first reason is the absence of test data for those "grade levels." The second reason is that all students tested at a given 1-12 grade level share the common experience of having been through the preceding grade levels down through grade level 1, but not all students have had early childhood and preschool experience.

The SKILLS data set contains 406,559 records housing the skills data for the following grade levels 1-12 skills:

Math.....2,441,431

Reading.....2,020,133
(actually reading
and oral language
combined)

TOTAL ---------
4,461,564

The contents of the records are identical to those of IDSHSK as shown in Appendix Figure C-9 of the ILLUSTRATIONS MANUAL.

The Derivation of IDSHTD (test data)

Step 1: From the entire MSRTS data base a subset of student records was selected. Each record selected contained test data for at least one CTBS or CAT test from the years 1983-1985. There were 351,011 such records selected.

Step 2: The 351,011 test records selected in step 3.1 were reduced by collapsing all tests (by test type) for a given student into a frequency count of outcomes of each test type (i.e., a frequency count for at or above grade and below grade math scores and a frequency count for at or above and below reading scores). In the foregoing data reduction, all CAT and CTBS math tests were classified as "math" and all CAT or CTBS reading, language and oral language tests were classified as "reading." The test records were then accumulated by grade level for each student. There were 144,148 "collapsed" test records produced by this reduction process. This collection of records is IDSHTD and is shown as Appendix Figure C-10 in the ILLUSTRATIONS MANUAL.

The Derivation of IDSHTDSK (skills and test data)

Step 1: For each of the 656,067 IDSHSK (skill data) records, two additional categories of skills for each skill type were created. The skills reported at each grade level are called "current" skills.

For each grade level in a student's record, the current skills were added to the current skills for the prior grade level. This computed measure is called "current + prior" skills.

For each grade level within a student's record, the number of skills reported at each prior grade level were added to the current skills. This computed measure is called "cumulative" skills.

Step 2: IDSHSK (the subset of skills data), the new categories of skills (see Step 1 above) and IDSHTD (the subset of the test data) were reduced jointly into IDSHTDSK as follows:

For the student number and grade level associated with each record in IDSHTD, a record having the same student number and grade level was sought in IDSHTDSK. When a match was found, the skill data and the test data were combined into a single record which, togegher with the student number and grade level, constituted a record IDSHTDSK. If no IDSHSK skill record match (for the IDSHTD test record) existed, the test data was combined with skill data set to zero for all "current" skill frequencies. Combined with student number and grade level, this record also constituted a member record for IDSHTDSK. This procedure led to the creation of 144,148 records in IDSHTDSK whose record content is listed below:

* student number
* grade level

* number of math skills (current)
* number of math skills (current + prior)
* number of math skills (cumulative)

* number of reading skills (current)
* number of reading skills (current + prior)
* number of reading skills (cumulative)

* number of math test scores at or above grade
* number of math test scores below grade

* number of reading test scores at or above grade
* number of reading test scores below grade.

See Appendix Figure C-11 in the ILLUSTRATION MANUAL.

D. SAS Analysis Module References

The SAS modules used to perform the analyses of the study may be referenced in two ways. Section 1 that follows cites an SAS module ID for each table and plot appearing in the ILLUSTRATIONS MANUAL. Section 2 presents a matrix map to the SAS modules used for the analysis of skills data.

SECTION 1. SAS Module Reference for Tables and Plots

A. EXPERIMENTAL SET A

Main effects of entry age, status, test type and grade level.

Table A1	SASTDGL3 7:39 Mon, Apr 28, 1986

B. EXPERIMENTAL SET B

Regression of scores at or above grade on entry age within status, test type and grade level.

Table B1	SASCGL2 7:49 Mon, Apr 28, 1986
Table B2	SASCGFR1 16:41 Thu, May 1, 1986
Table B3	SASTDGLM 16:24 Thu, May 1, 1986
Plot B1	SASCGGL2 7:49 Mon, Apr 28, 1986
Plot B2	SASCGFR1 16:41 Thu, May 1, 1986
Plot B3	SASTDGLM 8:49 Mon, May 5, 1986

C. EXPERIMENTAL SET C

Relationship between entry age and grade level.

Table C1,C3, Plot C1	SASCGMN1 11:35 Mon, May 5
Table C2,C4, Plot C2	SASCGMN2 15:03 Mon, May 5
Plot C3	Table C3
Plot C4	Table C4
Plot C5,C6,C7,C8,C9	Table C3,C4

D. EXPERIMENTAL SET D

Main effects of grade level, status and test type on test performance.

Table D1	SASTDGL6 11:39 Mon, Apr 28, 1986
Table D2	SASTDGL6 11:39 Mon, Apr 28, 1986

E. EXPERIMENTAL SET E

Regression of scores at or above grade on grade level within status, test type and entry age.

Table E1	SASCGGL3	7:55	Mon,	Apr	28,	1986
Table E2	SASCGFR2	16:51	Thu,	May	1,	1986
Table E3	SASTDG18	16:43	Mon,	May	5,	1986
Plot E1	SASCGGL3	7:55	Mon,	Apr	28,	1986
Plot E2	SASCGFR2	16:51	Thu,	May	1,	1986
Plot E3	SASTDG18	16:43	Mon,	May	5,	1986

F. EXPERIMENTAL SET F

A comparison, at each grade level, of migrant students having minimal exposure to the migrant program with migrant students who entered the migrant program at age six (6).

Table F1,F2,F3	SASEXGL8	11:27	Fri,	May	9,	1986
Table F4,F5,F6, Plot F1	SASEXGL5	8:50	Fri,	May	9,	1986
Table F7	SASEXGL7	10:03	Fri,	May	9,	1986

G. EXPERIMENTAL SET G

Joint effect of entry age and grade level on student performance.

Table G1,G2,G3,G4	SASTDG17	15:47	Mon,	May	5,	1986
Plot G1	SASTDGLM	8:49	Thu,	May	5,	1986

H. EXPERIMENTAL SET H

Descriptive statistics of math skills data in the SKILLS data set.

Table H1	SASSKGL4	18:35	Thu,	Jun	5,	1986
	SASSKGL3	8:46	Thu,	Jun	5,	1986
	SASTSG12	8:18	Thu,	Aug	7,	1986
	SASSKGL6	19:23	Thu,	Jun	5,	1986
	SASSKGL8	13:15	Thu,	Jul	24,	1986
	SASSKGL1	14:49	Mon,	May	12,	1986
Table H2	SASSKGL4	18:35	Thu,	Jun	5,	1986
Table H3	SASSKGL1	13:15	Thu,	Jul	24,	1986
Table H4	SASSKGL4	13:15	Thu,	Jul	24,	1986
Table H5	SASSKGL4	18:35	Thu,	Jun	5,	1986
Table H6	SASSKGL3	8:46	Mon,	May	12,	1986
Table H7	SASTSG12	8:18	Thu,	Aug	7,	1986
Table H8	SASSKGL4	19:23	Thu,	Jun	5,	1986
Table H9	SASSKGL1	14:49	Mon,	May	12,	1986
Table H10	SASSKGL8	13:15	Thu,	Jul	24,	1986
Table H11	SASSKGL7	10:56	Thu,	Jul	24,	1986
Table H12	SASSKGL8	13:15	Thu,	Jul	24,	1986
Table H13	SASSKGL8	13:15	Thu,	Jul	24,	1986
Plot H1	SASSKGL3	8:46	Mon,	May	12,	1986
Plot H2	SASSKGL8	13:15	Thu,	Jul	24,	1986
Plot H3	SASSKGL8	13:15	Thu,	Jul	24,	1986

I. EXPERIMENTAL SET I

Descriptive statistics of math topics represented by the math skills data in the SKILLS data set.

Table I1	SASSVFR2	2:42 Tue, May 6, 1986

J. EXPERIMENTAL SET J

Descriptive statistics of math skills data and math test data in IDSHTDSK. For clarity, the data set IDSHTDSK will be called "the skills/test data set."

Table J1	SASTSMN11	7:25 Thu, Jun 12, 1986
Table J2	SASTSGL3	16:01 Mon, Jul 14, 1986
	SASTSGL3	16:23 Mon, Jul 14, 1986
	SASTSG13	10:28 Thu, Aug 7, 1986
	SASTSG13	10:13 Thu, Aug 7, 1986
	SASTSGL3	17:14 Mon, Jul 14, 1986
	SASTSGL3	14:54 Mon, Jul 14, 1986
	SASTSGL2	11:34 Mon, Jul 14, 1986
	SASTSG10	15:56 Thu, Jul 24, 1986
	SASTSG11	8:07 Thu, Aug 7, 1986
	SASTSMN10	11:39 Fri, Jun 6, 1986
	SASTSM10	16:29 Tue, Jul 15, 1986
Table J3	SASTSGL4	15:43 Sat, Jul 19, 1986
	SASTSGL5	9:08 Mon, Jul 21, 1986
	SASTSGL6	16:00 Mon, Jul 21, 1986
	SASTSGL7	9:07 Mon, Jul 22, 1986
	SASTSGL8	15:21 Mon, Jul 21, 1986
	SASTSGL9	9:23 Tue, Jul 22, 1986
Table J4	SASTSM24	15:57 Wed, Aug 6, 1986
	SASTSM23	15:46 Wed, Aug 6, 1986
Plot J1	SASSMN10	11:39 Fri, Jun 6, 1986
	SASTSG10	15:56 Thu, Jul 24, 1986
	SASTSM10	16:29 Thu, Jul 17, 1986
Plot J2	SASTSMN10	14:25 Fri, Jun 6, 1986
Plot J3	SASTSGL7	9:07 Tue, Jul 22, 1986
Plot J4	SASTSGL8	15:51 Mon, Jul 21, 1986

K. EXPERIMENTAL SET K

Migrant student performance on standardized math tests as a function of varying numbers of math skills in MSRTS records. The data set used for this analysis was IDSHSK which, for clarity, will be called "the skills/test data set."

Table K1	SASTSMN1	17:42 Mon, May 19, 1986
Table K2	SASTSMN1	8:31 Mon, Jul 14, 1986
	SASTSMN1	17:15 Mon, May 5, 1986
Table K3	SASTSMN1	7:37 Fri, Jul 11, 1986

Table K4	SASTSMN1	16:57 Mon, May 19, 1986
Table K5	SASTSMN1	17:56 Mon, May 19, 1986
Plot K1	SASTSMN1	17:42 Mon, May 19, 1986
Plot K2	SASTSMN1	17:42 Mon, May 19, 1986
Plot K3	SASTSMN1	8:31 Mon, Jul 14, 1986
	SASTSMN1	17:15 Mon, May 5, 1986
Plot K4	SASTSMN1	8:31 Mon, Jul 14, 1986
Plot K5	SASTSMN1	17:15 Thu, May 29, 1986

L. EXPERIMENTAL SET L

Descriptive statistics of reading skills data in the SKILLS data set.

Table L1	SASSKGL4	18:35 Thu, Jun 5, 1986
	SASSKGL3	8:46 Thu, Jun 5, 1986
	SASTSG12	8:18 Thu, Aug 7, 1986
	SASSKGL6	19:23 Thu, Jun 5, 1986
	SASSKGL8	13:15 Thu, Jul 24, 1986
	SASSKGL1	14:49 Mon, May 12, 1986
Table L2	SASSKGL4	18:35 Thu, Jun 5, 1986
Table L3	SASSKGL1	13:15 Thu, Jul 24, 1986
Table L4	SASSKGL4	13:15 Thu, Jul 24, 1986
Table L5	SASSKGL4	18:35 Thu, Jun 5, 1986
Table L6	SASSKGL3	8:46 Mon, May 12, 1986
Table L7	SASTSG12	8:18 Thu, Aug 7, 1986
Table L8	SASSKGL4	19:23 Thu, Jun 5, 1986
Table L9	SASSKGL1	14:49 Mon, May 12, 1986
Table L10	SASSKGL8	13:15 Thu, Jul 24, 1986
Table L11	SASSKGL7	10:56 Thu, Jul 24, 1986
Table L12	SASSKGL8	13:15 Thu, Jul 24, 1986
Table L13	SASSKGL8	13:15 Thu, Jul 24, 1986
Plot L1	SASSKGL3	8:18 Thu, Aug 7, 1986
Plot L2	SASSKGL8	13:15 Thu, Jul 24, 1986
Plot L3	SASSKGL8	13:15 Thu, Jul 24, 1986

M. EXPERIMENTAL SET M

Descriptive statistics of reading levels represented by the reading skills data in the SKILLS data set.

Table M1	SASSVFR2	10:18 Tue, May 6, 1986
Plot M1	SASSVFR2	10:18 Tue, May 6, 1986

N. EXPERIMENTAL SET N

Descriptive statistics of reading skills data and reading test data in IDSHTDSK. For clarity, the data set IDSHTDSK will be called "the skills/test data set."

Table N1	SASTSMN11	7:25 Thu, Jun 12, 1986
Table N2	SASTSGL3	16:01 Mon, Jul 14, 1986
	SASTSGL3	16:23 Mon, Jul 14, 1986
	SASTSG13	10:28 Thu, Aug 7, 1986
	SASTSG13	10:13 Thu, Aug 7, 1986
	SASTSGL3	17:14 Mon, Jul 14, 1986
	SASTSGL3	14:54 Mon, Jul 14, 1986
	SASTSGL2	11:34 Mon, Jul 14, 1986
	SASTSG10	15:56 Thu, Jul 24, 1986
	SASTSG11	8:07 Thu, Aug 7, 1986
	SASTSMN10	11:39 Fri, Jun 6, 1986
	SASTSM10	16:29 Tue, Jul 15, 1986
Table N3	SASTSGL4	15:43 Sat, Jul 19, 1986
	SASTSGL5	9:08 Mon, Jul 21, 1986
	SASTSGL6	16:00 Mon, Jul 21, 1986
	SASTSGL7	9:07 Mon, Jul 22, 1986
	SASTSGL8	15:21 Mon, Jul 21, 1986
	SASTSGL9	9:23 Tue, Jul 22, 1986
Table N4	SASTSM24	15:57 Wed, Aug 6, 1986
	SASTSM23	15:46 Wed, Aug 6, 1986
Plot N1	SASSMN10	11:39 Fri, Jun 6, 1986
	SASTSG10	15:56 Thu, Jul 24, 1986
	SASTSM10	16:29 Thu, Jul 17, 1986
Plot N2	SASTSMN10	14:25 Fri, Jun 6, 1986
Plot N3	SASTSGL7	9:07 Tue, Jul 22, 1986
Plot N4	SASTSGL8	15:51 Mon, Jul 21, 1986

O. EXPERIMENTAL SET O

Migrant student performance on standardized reading tests as a function of varying numbers of reading skills in MSRTS records. The data set used for this analysis was IDSHSK which, for clarity, will be called "the skills/test data set."

Table O1	SASTSMN2	16:57 Mon, May 19, 1986
Table O2	SASTSMN2	8:09 Mon, Jul 14, 1986
Table O3	SASTSMN2	8:18 Mon, Jul 14, 1986
Table O4	SASTSMN2	17:20 Mon, May 19, 1986
Table O5	SASTSMN2	17:42 Mon, May 19, 1986
Table O6	SASTAMN2	16:57 Mon, May 19, 1986
Plot O1	SASTSMN2	16:57 Mon, May 19, 1986
Plot O2	SASTSMN2	8:09 Mon, Jul 14, 1986

SECTION 2. SAS Modules used for the Analyses of Skills Data

Map 1 Analysis of variance modules for
NUMBER OF SKILLS PER STUDENT = GRADE LEVEL

Map 2 Analysis of variance modules for
NUMBER OF TESTS PER STUDENT = GRADE LEVEL
NUMBER OF TESTS PER STUDENT = NUMBER OF SKILLS PER STUDENT

Map 3 Linear Regression for
NUMBER OF SKILLS PER STUDENT = GRADE LEVEL

Map 4 Linear Regression for
NUMBER OF TESTS PER STUDENT = GRADE LEVEL
NUMBER OF TESTS PER STUDENT = NUMBER OF SKILLS PER STUDENT

Map 5 Linear regression for
PERCENT AT/ABOVE GRADE = NUMBER OF SKILLS PER STUDENT

Map 6 Polynomial Regression for
NUMBER OF SKILLS PER STUDENT = GRADE LEVEL

Map 7 Polynomial Regression for
NUMBER OF TESTS PER STUDENT = GRADE LEVEL
NUMBER OF TESTS PER STUDENT = NUMBER OF SKILLS PER STUDENT

Map 8 Polynomial regression for
PERCENT AT/ABOVE GRADE = NUMBER OF SKILLS PER STUDENT

Certain symbols are used as abbreviations on the maps follows:

_C = skills example MATH_C means math skills
_T = tests example READ_T means reading tests

MAP 1

ANALYSIS OF VARIANCE
POPULATION ANALYSIS MATRIX FOR NUMBER OF SKILLS

(shsk in lower case. SHTDSK IN UPPER CASE)

POPULATION	SAS MODEL: # SKILLS = GRADE LEVEL SAS MODULE ID
MATH_C	TSGL3 16:01 7/14 TSGL2 12:47 7/14 skgl4 18:35 6/05
READ_C	TSGL3 16:01 7/14 TSGL2 12:47 7/14 skgl4 18:35 6/05
MATH_T	TSGL3 16:23 7/14
READ_T	TSGL3 16:23 7/14
MATH_C AND READ_C	TSG13 10:28 8/7 skgl7 10:56 7/24
MATH_C OR READ_C	TSG13 10:28 8/7 skgl6 19:23 6/03
MATH_T AND READ_T	TSG13 10:13 8/7
MATH_T OR READ_T	TSGL3 17:14 7/14
MATH_C AND MATH_T	TSGL3 14:54 7/14 TSGL2 14:32 7/14
READ_C AND READ_T	TSGL2 14:54 7/14 TSGL2 14:54 7/14
(MATH_C OR READ_C) AND (MATH_T)	TSGL2 11:34 7/14
(MATH_C OR READ_C) AND (READ_T)	TSGL2 11:34 7/14
MATH_I AND MATH_T	NA-3
READ_I AND READ_T	NA-3

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 2

ANALYSIS OF VARIANCE
POPULATION ANALYSIS MATRIX FOR NUMBER OF TESTS

POPULATION	SAS MODEL: # TESTS = SKILLS INTERVAL SAS MODULE ID	# TESTS = GRADE LEVEL SAS MODULE ID
MATH_C	NA-2	TSGL4 15:43 7/19
READ-C	NA-2	TSGL4 15:43 7/19
MATH_T	TSM24 15:57 8/6	TSGL4 15:43 7/19
READ_T	TSM24 15:57 8/6	TSGL4 15:43 7/19
MATH_C AND READ_C	NA-2	TSGL5 9:08 7/21
MATH_C OR READ_C	NA-2	TSGL5 9:08 7/21
MATH_T AND READ_T	TSM24 15:57 8/6	TSGL5 9:08 7/21
MATH_T OR READ_T	TSM24 15:57 8/6	TSGL5 9:08 7/21
MATH_C AND MATH_T	NA-2	TSGL5 9:08 7/21
READ_C AND READ_T	NA-2	TSGL5 9:08 7/21
(MATH_C OR READ_C) AND (MATH_T)	NA-2	TSGL4 15:43 7/19
(MATH_C OR READ_C) AND (READ_T)	NA-2	TSGL4 15:43 7/19
MATH_I AND MATH_T	TSM24 15:57 8/6	NA-3
READ_I AND READ_T	TSM24 15:17 8/6	NA-3
MATH_I AND READ_I	TSM24 15:57 8/6	NA-3
MATH_I OR READ_I	TSM24 15:57 8/6	NA-3
(MATH_I OR READ_I) AND MATH_T	TSM24 15:57 8/6	NA-3
(MATH_I OR READ_I) AND READ_T	TSM24 15:57 8/6	

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 3

LINEAR REGRESSION
POPULATION ANALYSIS MATRIX FOR NUMBER OF SKILLS

(shsk in lower case. SHTDSK IN UPPER CASE.)

POPULATION	SAS MODEL: # SKILLS = GRADE LEVEL SAS MODULE ID
MATH_C	TSGL3 16:01 7/14
	TSGL2 12:47 7/14
	skgl3 8:46 5/12
READ-C	TSGL3 16:01 7/14
	TSGL2 12:47 7/14
	skgl3 8:46 5/12
MATH_T	TSGL3 16:23 7/14
	SMN10 11:39 6/06
READ_T	TSGL3 16:23 7/14
	SMN10 11:39 6/06
MATH_C AND READ_C	TSG10 15:56 7/24
	skgl8 13:15 7/24
MATH_C OR READ_C	TSG10 15:56 7/24
	skgl1 14:49 5/12
	skgl8 13:15 7/24
MATH_T AND READ_T	TSG10 15:56 7/24
MATH_T OR READ_T	TSGL3 17:14 7/14
	TSG10 15:56 7/24
MATH_C AND MATH_T	TSGL3 14:54 7/14
	SMN10 11:39 6/06
	TSGL2 14:32 7/14
	TSM10 16:29 7/15
READ_C AND READ_T	TSGL3 14:54 7/14
	SMN10 11:39 6/06
	TSGL2 14:32 7/14
	TSM10 16:29 7/15
(MATH_C OR READ_C) AND (MATH_T)	TSM10 16:29 7/15 TSGL2 11:34 7/14
(MATH_C OR READ_C) AND (READ_T)	TSM10 16:29 7/15 TSGL2 11:34 7/14

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 4

LINEAR REGRESSION
POPULATION ANALYSIS MATRIX FOR NUMBER OF TESTS

POPULATION	SAS MODEL: # TESTS = SKILLS INTERVAL SAS MODULE ID	# TESTS = GRADE LEVEL SAS MODULE ID
MATH_C	NA-2	TSGL6 16:00 7/21
READ-C	NA-2	TSGL6 16:00 7/21
MATH_T	TSM23 15:46 8/6	TSGL6 16:00 7/21
READ_T	TSM23 15:46 8/6	TSGL6 16:00 7/21
MATH_C AND READ_C	NA-2	TSGL7 9:07 7/22
MATH_C OR READ_C	NA-2	TSGL7 9:07 7/22
MATH_T AND READ_T	TSM23 15:46 8/6	TSGL7 9:07 7/22
MATH_T OR READ_T	TSM23 15:46 8/6	TSGL7 9:07 7/22
MATH_C AND MATH_T	NA-2	TSGL7 9:07 7/22
READ_C AND READ_T	NA-2	TSGL7 9:07 7/22
(MATH_C OR READ_C) AND (MATH_T)	NA-2	TSGL6 16:00 7/21
(MATH_C OR READ_C) AND (READ_T)	NA-2	TSGL6 16:00 7/21
MATH_I AND READ_I	TSM23 15:46 8/6	NA-3
MATH_I OR READ_I	TSM23 15:46 8/6	NA-3
MATH_I AND MATH_T	TSM23 15:46 8/6	NA-3
READ_I AND READ_T	TSM23 15:46 8/6	NA-3
(MATH_I OR READ_I) AND (MATH_T)	TSM23 15:46 8/6	NA-3
(MATH_C OR READ_C) AND (READ_T)	TSM23 15:46 8/6	NA-3

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 5

LINEAR REGRESSION
POPULATION ANALYSIS MATRIX FOR TEST SCORES

POPULATION	SAS MODEL: BY GRADE; PCT = SKILL_INTERVAL SAS MODULE ID
MATH_C	NA-2
READ-C	NA-2
MATH_T	TSMN1 17:42 5/19
READ_T	TSMN2 16:57 5/19
MATH_C AND READ_C	NA-3
MATH_C OR READ_C	NA-3
MATH_T AND READ_T	TSM22 15:22 7/23
MATH_T OR READ_T	TSM22 15:22 7/23
MATH_C AND MATH_T	NA-2
READ_C AND READ_T	NA-2
(MATH_C OR READ_C) AND (MATH_T)	NA-2
(MATH_C OR READ_C) AND (READ_T)	NA-2
MATH_I AND READ_I	TSM22 15:22 7/23
MATH_I OR READ_I	TSM22 15:22 7/23
MATH_I AND MATH_T	TSMN1 8:31 7/14 TSMN1 17:15 5/29
READ_I AND READ_T	TSMN2 8:09 7/14
(MATH_I OR READ_I) AND (MATH_T)	TSMN1 7:37 7/11
(MATH_I OR READ_I) AND (READ_T)	TSMN2 8:18 7/14

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 6

POLYNOMIAL REGRESSION
POPULATION ANALYSIS MATRIX FOR NUMBER OF SKILLS

(shsk in lower case. SHTDSK IN UPPER CASE.)

POPULATION	SAS MODEL: # SKILLS = GRADE LEVEL SAS MODULE ID
MATH_C	TSG11 8:07 8/7 tsg12 8:18 8/7
READ-C	TSG11 8:07 8/7 tsg12 8:18 8/7
MATH_T	SMN10 11:39 6/06
READ_T	SMN10 11:39 6/06
MATH_C AND READ_C	TSG10 15:56 7/24 skg18 13:15 7/24
MATH_C OR READ_C	TSG10 15:56 7/24 skg18 13:15 7/24
MATH_T AND READ_T	TSG10 15:56 7/24
MATH_T OR READ_T	TSG10 15:56 7/24
MATH_C AND MATH_T	SMN10 11:39 6/06 TSM10 16:29 7/15
READ_C AND READ_T	SMN10 11:39 6/06 TSM10 16:29 7/15
(MATH_C OR READ_C) AND (MATH_T)	TSM10 16:29 7/15
(MATH_C OR READ_C) AND (READ_T)	TSM10 16:29 7/15

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 7

POLYNOMIAL REGRESSION
POPULATION ANALYSIS MATRIX FOR NUMBER OF TESTS

POPULATION	SAS MODEL: # TESTS = SKILLS INTERVAL SAS MODULE ID	# TESTS = GRADE LEVEL SAS MODULE ID
MATH_C	NA-2	TSGL8 15:21 7/21
READ-C	NA-2	TSGL8 15:21 7/21
MATH_T	TSM23 15:46 8/6	TSGL8 15:21 7/21
READ_T	TSM23 15:46 8/6	TSGL8 15:21 7/21
MATH_C AND READ_C	NA-2	TSGL9 9:23 7/22
MATH_C OR READ_C	NA-2	TSGL9 9:23 7/22
MATH_T AND READ_T	TSM23 15:46 8/6	TSGL9 9:23 7/22
MATH_T OR READ_T	TSM23 15:46 8/6	TSGL9 9:23 7/22
MATH_C AND MATH_T	NA-2	TSGL9 9:23 7/22
READ_C AND READ_T	NA-2	TSGL9 9:23 7/22
(MATH_C OR READ_C) AND (MATH_T)	NA-2	TSGL8 15:21 7/21
(MATH_C OR READ_C) AND (READ_T)	NA-2	TSGL8 15:21 7/21
MATH_I AND READ_I	TSM23 15:46 8/6	NA-3
MATH_I OR READ_I	TSM23 15:46 8/6	NA-3
MATH_I AND MATH_T	TSM23 15:46 8/6	NA-3
READ_I AND READ_T	TSM23 15:46 8/6	NA-3
(MATH_I OR READ_I) AND (MATH_T)	TSM23 15:46 8/6	NA-3
(MATH_C OR READ_C) AND (READ_T)	TSM23 15:46 8/6	NA-3

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

MAP 8

POLYNOMIAL REGRESSION
POPULATION ANALYSIS MATRIX FOR TEST SCORES

POPULATION	SAS MODEL: BY GRADE; PCT = SKILL_INTERVAL SAS MODULE ID
MATH_C	NA-2
READ-C	NA-2
MATH_T	TSMN1 17:42 5/19
READ_T	TSMN2 16:57 5/19
MATH_C AND READ_C	NA-3
MATH_C OR READ_C	NA-3
MATH_T AND READ_T	TSM22 15:22 7/23
MATH_T OR READ_T	TSM22 15:22 7/23
MATH_C AND MATH_T	NA-2
READ_C AND READ_T	NA-2
(MATH_C OR READ_C) AND (MATH_T)	NA-2
(MATH_C OR READ_C) AND (READ_T)	NA-2
MATH_I AND READ_I	TSM22 15:22 7/23
MATH_I OR READ_I	TSM22 15:22 7/23
MATH_I AND MATH_T	TSMN1 8:31 7/14 TSMN1 17:15 5/29
READ_I AND READ_T	TSMN2 8:09 7/14
(MATH_I OR READ_I) AND (MATH_T)	TSMN1 7:37 7/11
(MATH_I OR READ_I) AND (READ_T)	TSMN2 8:18 7/14

NA-1 Must have test data in each record selected to perform the analysis indicated.
NA-2 Skills data must be in intervals to perform the analysis indicated.
NA-3 Skills data cannot be in interval form to perform the analysis indicated.

E. GLOSSARY OF TERMS

Student Performance

Unless otherwise stated, this term means the percent of a group whose test scores are at/above grade level.

Entry Age

The age of a migrant student when first enrolled into a migrant education program AS REPORTED TO THE MSRTS.

Exposure to the Migrant Education Program

The length of time elapsed between a student's first enrollment in a school as a migrant student and the experimental cutoff time being used for comparison purposes. The quantity exposure is a unit of time only and implies neither intensity nor continuity of exposure nor quality of the program to which the student is exposed. Exposure may be an inferred quantity such as when groups of different entry ages are compared.

Pseudo Hypotheses

Each of the major independent variables investigated in this study is discussed as a "Pseudo Hypotheses." Actual hypotheses used for decision rules, though technically correct, reveal little of the investigative intent at hand. In the following, each Pseudo Hypothesis (abbreviated PH) represents a liberalized statement of a hypothesis regarding the effects of the independent variable under discussion. Although a Pseudo Hypothesis may be directional, its technical counterpart need not be.

Experimental Sets

The statistical analyses were performed in sets. Each set investigated the effects of the independent variable of principal interest (and its relationship to the other independent variables) on a particular population. Each separate analysis is referred to as an experiment. Thus each experimental set is a series of experiments whose primary purpose is to increase understanding of the principal variable and whose secondary purpose is to identify other contributing variables.

SOME STATISTICAL TERMS

MEAN SQUARE (MODEL) = MODEL SS DIVIDED BY MODEL df.
MEAN SQUARE (ERROR) = ERROR SS DIVIDED BY ERROR df.

MEAN SQUARE (ERROR) (abbreviated MSE) is an estimate of sigma squared.
Square root of MSE = ROOT MEAN SQ.
ERROR(MSE) = sigma.

R-SQUARE = MODEL SUM OF SQUARES DIVIDED BY CORRECTED SUM OF SQUARES.

COEFFICIENT OF VARIATION (C.V.) = (ROOT MSE DIVIDED BY MEAN OF DEPENDENT VARIABLE) X 100.

F = MEAN SQUARE (MODEL) DIVIDED BY MEAN SQUARE (ERROR).

AN INTRA POPULATION ANALYSIS OF MIGRANT STUDENT PERFORMANCE BASED ON THE MIGRANT STUDENT RECORD TRANSFER SYSTEM DATA BASE

ILLUSTRATIONS MANUAL

by
Eugene deG. Hackett
James V. McManus

August 25, 1986

ILLUSTRATIONS MANUAL

TABLE OF CONTENTS

A. EXPERIMENTAL SET A

Main effects of entry age, status, test type and grade level.

PAGE

Table A1 1
GLM ANOVA table for the analysis of student performance as a function of status, test type, grade level and entry age.

- -

B. EXPERIMENTAL SET B

Regression of scores at or above grade on entry age within; status, test type and grade level.

PAGE

Table B1 2
Data from the analysis of the regression of student performance on entry age for each status.

Table B2 3
Data from the analysis of the regression of student performance on entry age for each test type.

Table B3 5
Synopsis of data from the analysis of regression of student performance on entry age for each grade level.

Plot B1 6
Regression of student performance on entry age for each status.

Plot B2 7
Regression of student performance on entry age for each test type.

Plot B3 8
Regression of student performance on entry age for each of several grade levels.

- -

C. EXPERIMENTAL SET C

Relationship between entry age and grade level.

	PAGE
Table C1 Regression table of mean entry age on grade level.	9
Table C2 Regression table of mean grade level on entry age.	9
Table C3 Distribution statistics of entry age for each grade level.	10
Table C4 Distribution statistics of grade level for each entry age.	11
Plot C1 Regression of mean entry age on grade level.	12
Plot C2 Regression of mean grade level on entry age.	13
Plot C3 Range of (90%) of entry ages for each grade level.	14
Plot C4 Range of (90%) of grade levels for each entry age.	15
Plot C5 Overlap of entry age distributions and grade level distributions.	16
Plot C6 Percent of overlap between two entry age/grade level distributions.	17
Plot C7 Percent of non-overlap between two entry age/grade level distributions.	18
Plot C8 Coefficient of variation for the entry age/grade level distributions.	19
Plot C9 Means and standard deviations for the entry age/grade level distributions.	20

D. EXPERIMENTAL SET D

Main effects of grade level, status and test type on test performance.

PAGE

Table D1 21
GLM ANOVA table for the analysis of student performance as a function of status, test type, grade level and interactions thereof.

Table D2 22
Table of means of the main effect variables.

- -

E. EXPERIMENTAL SET E

Regression of scores at or above grade on grade level within; status, test type and entry age.

PAGE

Table E1 23
Data from the analysis of the regression of student performance on grade level for each status.

Table E2 24
Data from the analysis of the regression of student performance on grade level for each test type.

Table E3 26
Synopsis of data from the analysis of regression of student performance on grade level for each entry age.

Plot E1 27
Regression of student performance on grade level for each status.

Plot E2 28
Regression of student performance on grade level for each test type.

Plot E3 29
Regression of student performance on grade level for each (even numbered) entry age.

- -

F. EXPERIMENTAL SET F

A comparison, at each grade level, of migrant students having minimal exposure to the migrant program with migrant students who entered the migrant program at age six (6).

PAGE

Table F1 30
List of variables and levels making up ANOVA.

Table F2 30
GLM ANOVA table for the analysis of student performance as a function of experimental group, grade level, test type, status and interactions thereof.

Table F3 31
List of means of the main effect variables.

Table F4 32
Data used to analyze the regression of the percent of scores at/or above grade on experimental groups 1 and 6.

Table F5 32
Mean percent of scores at/or above grade level for each of two experimental groups by grade level.

Table F6 33
Data from the analysis of the regression of student performance on each of two experimental groups by grade level.

Table F7 34
Data from the GLM analysis of difference in mean ages of two experimental groups.

Plot F1 35
Regression of student performance on grade level for each of two experimental groups.

- -

G. EXPERIMENTAL SET G

Joint effect of entry age and grade level on student performance.

PAGE

Table G1 36
Data from a simple multiple (regression) analysis of the regression of student performance on entry age and grade level.

Table G2 36
Data from a polynomial multiple (regression) analysis of the regression of student performance on entry age and grade level and entry age X grade level.

Table G3 37
Data from a full model polynomial multiple (regression) analysis of the regression of student performance on entry age and grade level and grade level X grade level and entry age X entry age and entry age X grade level.

Table G4 37
Data from a (partial model) polynomial multiple (regression) analysis of the regression of student performance on grade level and grade level X grade level and entry age X entry age.

Plot G1 38
Performance as a function of entry age and grade level.

H. EXPERIMENTAL SET H

Descriptive statistics of math skills data in the SKILLS data set.

PAGE

Table H1 39
Summary analysis table for the analysis of the mean number of math skills as a function of grade level in the SKILLS data set.

Table H2 39
Table of mean number of math skills per student by grade level for the population of all records having at least one math skill.

Table H3 40
Table of mean number of math skills per student by grade level for the population of all records having at least one math skill or reading skill or both.

PAGE

Table H4 40
Table of mean number of math skills per student by grade level for the population of all records having at least one math skill and one reading skill.

Table H5 41
GLM ANOVA of number of math skills X grade level for the population of records from the SKILLS data set having at least one math skill.

Table H6 41
GLM linear regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill.

Table H7 41
GLM polynomial regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill.

Table H8 42
GLM ANOVA of number of math skills X grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Table H9 42
GLM linear regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Table H10 42
GLM polynomial regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Table H11 43
GLM ANOVA of number of math skills X grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

Table H12 43
GLM linear regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

PAGE

Table H13 43
GLM polynomial regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

Plot H1 44
GLM linear regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill.

Plot H2 45
GLM polynomial regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Plot H3 46
GLM linear regression of mean number of math skills per student on grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

- -

I. EXPERIMENTAL SET I

Descriptive statistics of math topics represented by the math skills data in the SKILLS data set.

PAGE

Table I1 47
A two way table of the frequency and percent of math skills by topical area reported for each grade level.

- -

J. EXPERIMENTAL SET J

Descriptive statistics of math skills data and math test data in IDSHTDSK. For clarity, the data set IDSHTDSK will be called "the skills/test data set."

PAGE

Table J1 48
Summary parameters of the SKILLS data set and the SKILLS/TEST data set.

Table J2 49
GLM ANOVA and regression analysis summary table for the mean number of math skills per student as a function of grade level.

PAGE

Table J3 49
GLM ANOVA and regression analysis summary table for the mean number of math tests per student as a function of grade level.

Table J4 50
GLM ANOVA and regression analysis summary table for the mean number of math tests per student as a function of math skills per student.

Plot J1 51
Polynomial regression of the mean number of math skills per student on grade level for five populations.

Plot J2 52
Best fit polynomial of mean number of current, current + previous and cumulative math skills per student.

Plot J3 53
Linear regression of mean number of math tests on grade level for the population of records having both math and reading tests.

Plot J4 54
Polynomial regression of mean number of math tests on grade level for the population of records having math skills.

- -

K. EXPERIMENTAL SET K

Migrant student performance on standardized math tests as a function of varying numbers of math skills in MSRTS records. The data set used for this analysis was IDSHSK which for clarity will be called "the skills/test data set."

PAGE

Table K1 55
Summary table of analysis for regression of math test performance on the number of math skills in the population of records containing math tests, but not necessarily math skills.

Table K2 55
Summary table of analysis for regression of math test performance on the number of math skills in the population of records containing math tests and math skills.

Table K3 55
Summary table of analysis for regression of math test performance on the number of math skills in the population of records containing math skills or reading skills or both and math tests.

PAGE

Table K4 56
Summary table of analysis for regression of math test performance on the number of math skills in the population of records containing math tests and possibly current + previous math skills.

Table K5 56
Summary table of analysis for regression of math test performance on the number of math skills in the population of records containing math tests and possibly cumulative math skills.

Plot K1 57
GLM linear regression curves of best fit of percent of math scores at or above grade level on the number of current math skills in student records having math test data, but not necessarily math skills data.

Plot K2 58
GLM polynomial regression curves of best fit of percent of math scores at or above grade level on the number of current math skills in student records having math test data, but not necessarily math skills data.

Plot K3 59
GLM linear regression curves of best fit of percent of math scores at or above grade level on the number of current math skills in student records having math test data and math skills data.

Plot K4 60
GLM polynomial regression curves of best fit of percent of math scores at or above grade level on the number of current math skills in student records having math test data and math skills data.

Plot K5 61
GLM linear regression of percent of math scores at or above grade level on the number of current math skills in student records having math test data and math skills data.

- -

L. EXPERIMENTAL SET L

Descriptive statistics of reading skills data in the SKILLS data set.

PAGE

Table L1 62
Summary analysis table for the analysis of the mean number of reading skills as a function of grade level in the SKILLS data set.

PAGE

Table L2 62
Table of mean number of reading skills per student by grade level for the population of all records having at least one reading skill.

Table L3 63
Table of mean number of reading skills per student by grade level for the population of all records having at least one math skill or reading skill or both.

Table L4 63
Table of mean number of reading skills per student by grade level for the population of all records having at least one math skill and one reading skill.

Table L5 64
GLM ANOVA of number of reading skills X grade level for the population of records from the SKILLS data set having at least one reading skill.

Table L6 64
GLM linear regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one reading skill.

Table L7 64
GLM polynomial regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one reading skill.

Table L8 65
GLM ANOVA of number of reading skills X grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Table L9 65
GLM linear regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Table L10 65
GLM polynomial regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Table L11 66
GLM ANOVA of number of reading skills X grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

PAGE

Table L12 66
GLM linear regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

Table L13 66
GLM polynomial regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

Plot L1 67
GLM polynomial regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one reading skill.

Plot L2 68
GLM polynomial regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one math skill or one reading skill or both.

Plot L3 69
GLM polynomial regression of mean number of reading skills per student on grade level for the population of records from the SKILLS data set having at least one math skill and one reading skill.

- - - -

N. EXPERIMENTAL SET N

Descriptive statistics of reading skills data and reading test data in IDSHTDSK. For clarity, the data set IDSHTDSK will be called "the skills/test data set."

PAGE

Table N1 73
Summary parameters of the SKILLS data set and the SKILLS/TEST data set.

Table N2 74
GLM ANOVA and regression analysis summary table for the mean number of reading skills per student as a function of grade level.

Table N3 74
GLM ANOVA and regression analysis summary table for the mean number of reading tests per student as a function of grade level.

Table N4 75
GLM ANOVA and regression analysis summary table for the mean number of reading tests per student as a function of reading skills per student.

Plot N1 76
Best fit curve for the regression of the mean number of reading skills per student on grade level for four populations.

Plot N2 77
Best fit polynomial of mean number of current, current + previous and cumulative reading skills per student.

Plot N3 78
Best fit curve for the regression of mean number of reading tests per student on grade level for three populations.

Plot N4 79
Polynomial regression of mean number of reading tests on the number of skills per student for the population of records having reading tests.

- -

O. EXPERIMENTAL SET O

Migrant student performance on standardized reading tests as a function of varying numbers of reading skills in MSRTS records. The data set used for this analysis was IDSHSK which for clarity will be called "the skills/test data set."

PAGE

Table O1 80
Summary table of analysis for regression of test performance on the number of reading skills in the population of records containing reading tests, but not necessarily reading skills.

Table O2 80
Summary table of analysis for regression of test performance on the number of reading skills in the population of records containing reading tests and reading skills.

Table O3 80
Summary table of analysis for regression of test performance on the number of reading skills in the population of records containing math skills or reading skills or both and reading tests.

Table O4 81
Summary table of analysis for regression of reading test performance on the number of reading skills in the population of records containing reading tests and possibly current + previous reading skills.

Table O5 81
Summary table of analysis for regression of reading test performance on the number of reading skills in the population of records containing reading tests and possibly cumulative reading skills.

Table O6 82
Frequency table of reading skills intervals by grade level.

Plot O1 83
GLM regression curves of best fit of percent of reading scores at or above grade level on the number of current reading skills in student records having reading test data, but not necessarily reading skills data.

Plot O2 84
GLM regression curves of best fit of percent of reading scores at or above grade level on the number of current reading skills in student records having reading test data and reading skills data.

ILLUSTRATIONS MANUAL

APPENDIX

PAGE

Appendix Figure C-1 85
Data sets selected from the MSRTS data base for the analysis of populations of skills and test data.

Appendix Figure C-2 86
Four populations selected from data set skills.

Appendix Figure C-3 87
Twelve populations selected from IDSHTDSK.

Appendix Figure C-4 88
Number of records making up the populations selected from the SKILLS and IDSHTDSK data sets.

Appendix Figure C-5 89
Number of records and skills in populations selected for the SKILLS data set.

Appendix Figure C-6 89
Number of records, skills and tests in populations selected from the IDSHTDSK data set.

Appendix Figure C-7 90
Number of records making up the mutually exclusive populations selected from the SKILLS data sets.

Appendix Figure C-8 90
Number of records making up the mutually exclusive populations selected from the IDSHTDSK data sets.

Appendix Figure C-9 91
Fields making up a record in IDSHSK.

Appendix Figure C-10 91
Fields making up a record in IDSHTD.

Appendix Figure C-11 92
Fields making up a record in IDSHTDSK.

GENERAL LINEAR MODELS PROCEDURE SASTDGL3

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
STATUS	3	1 2 3
TEST TYP	5	101 102 103 301 302
GRADE	12	1 2 3 4 5 6 7 8 9 10 11 12
ENTR AGE	22	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

NUMBER OF OBSERVATIONS IN DATA SET = 2859

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 2541 OBSERVATIONS CAN BE USED IN THIS ANALYSIS.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: AT ABOVE % AT OR ABOVE GRADE LEVEL
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	38	10098.22456047	265.74275159	589.44	0.0001	0.899521	151.6825
ERROR	2502	1127.99545251	0.45083751		ROOT MSE		AT ABOVE MEAN
CORRECTED TOTAL	2540	11226.22001298			0.67144435		0.44266424

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
STATUS	2	118.10593623	130.99	0.0001	2	76.30113649	84.62	0.0001
TEST TYP	4	3451.61651012	1914.00	0.0001	4	3562.91538661	1975.72	0.0001
GRADE	11	6179.82622161	1246.13	0.0001	11	3718.79857955	749.88	0.0001
ENTR AGE	21	348.67589251	36.83	0.0001	21	348.67589251	36.83	0.0001

TABLE A1

SASCGGL2

STATUS=1

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	3719949.46801001	3719949.46801001	664.92	0.0001	0.972219	182.9877
ERROR	19	106296.31218099	5594.54274637		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	20	3826245.78019100			74.79667604		40.87523961

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	3719949.46801001	664.92	0.0001	1	3719949.46801001	664.92	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	62.26493844	104.26	0.0001	0.59722324
ENTR AGE	-2.17744173	-25.79	0.0001	0.08444235

STATUS=2

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	4682737.73278915	4682737.73278915	345.98	0.0001	0.950546	276.6693
ERROR	18	243626.00395019	13534.77799723		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	19	4926363.73673934			116.33906479		42.04986224

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	4682737.73278915	345.98	0.0001	1	4682737.73278915	345.98	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	67.31532322	60.70	0.0001	1.10893867
ENTR AGE	-2.69800268	-18.60	0.0001	0.14505010

STATUS=3

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	18717551.08260470	18717551.08260470	875.27	0.0001	0.977660	343.0361
ERROR	20	427695.44027852	21384.77201393		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	21	19145246.52288320			146.23533094		42.62972443

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	18717551.08260470	875.27	0.0001	1	18717551.08260470	875.27	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	73.18808830	96.34	0.0001	0.75969760
ENTR AGE	-2.66962609	-29.59	0.0001	0.09699584

2

TABLE B1

TEST TYP=101

CTBS READING

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	7202440.39916200	7202440.39916200	377.06	0.0001	0.954438	412.4039
ERROR	18	343824.98153764	19101.38786320		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	19	7546265.38069964			138.20777063		33.51272159

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	7202440.39916200	377.06	0.0001	1	7202440.39916200	377.06	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	60.42256576	54.21	0.0001	1.11460027
ENTR AGE	-2.90098947	-19.42	0.0001	0.14939593

TEST TYP=102

CTBS ARITHMETIC

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	5790599.82329079	5790599.82329079	540.79	0.0001	0.966059	197.8470
ERROR	19	203445.29198096	10707.64694637		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	20	5994045.11527175			103.47776064		52.30191570

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	5790599.82329079	540.79	0.0001	1	5790599.82329079	540.79	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	79.33470386	92.76	0.0001	0.85527757
ENTR AGE	-2.68406288	-23.25	0.0001	0.11541917

TEST TYP=103

CTBS LANGUAGE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	5645150.15739487	5645150.15739487	510.30	0.0001	0.965928	276.8548
ERROR	18	199123.27709999	11062.40428333		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	19	5844273.43449486			105.17796482		37.99029564

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	5645150.15739487	510.30	0.0001	1	5645150.15739487	510.30	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	65.69437723	66.95	0.0001	0.98126744
ENTR AGE	-2.96741812	-22.59	0.0001	0.13136079

TABLE B2 (cont'd next page)

TEST TYP=301

CAT READING

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	4932001.84350202	4932001.84350202	327.15	0.0001	0.947849	331.8306
ERROR	18	271361.78921251	15075.65495625		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	19	5203363.63271454			122.78295874		37.00169355

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	4932001.84350202	327.15	0.0001	1	4932001.84350202	327.15	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	63.04488857	54.84	0.0001	1.14959679
ENTR AGE	-2.59264386	-18.09	0.0001	0.14334064

TEST TYP=302

CAT ARITHMETIC

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	3405605.29939057	3405605.29939057	108.69	0.0001	0.851198	310.9213
ERROR	19	595348.89609337	31334.15242597		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	20	4000954.19548394			177.01455428		56.93227716

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	3405605.29939057	108.69	0.0001	1	3405605.29939057	108.69	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	82.12227800	47.70	0.0001	1.72162034
ENTR AGE	-2.22051635	-10.43	0.0001	0.21299323

TABLE B2 (cont'd)

Synopsis of data from the analysis of regression of student performance on Entry age in each Grade level

Grade	Mean	Intcpt	Slope	R-Sq	C.V.	PR > F
01	78.71	78.59	0.85	.52	116.90	.0126
02	65.22	67.00	-0.43	.32	115.01	.0446
03	54.15	59.76	-0.80	.77	104.13	.0001
04	45.36	56.82	-1.05	.72	200.49	.0001
05	43.40	55.81	-1.36	.81	209.40	.0001
06	38.06	51.26	-1.10	.67	288.74	.0001
07	40.04	49.70	-0.80	.55	271.63	.0007
08	40.09	51.82	-0.99	.55	294.21	.0003
09	33.94	48.72	-1.23	.75	263.19	.0001
10	36.91	46.45	-1.13	.86	134.48	.0001
11	34.34	45.81	-1.28	.76	203.07	.0001
12	29.91	40.57	-1.00	.64	167.31	.0001

Mean=Mean percent of test scores at or above grade level across all entry ages

R-Sq=R-Square=Mean square / corrected total sum of squares

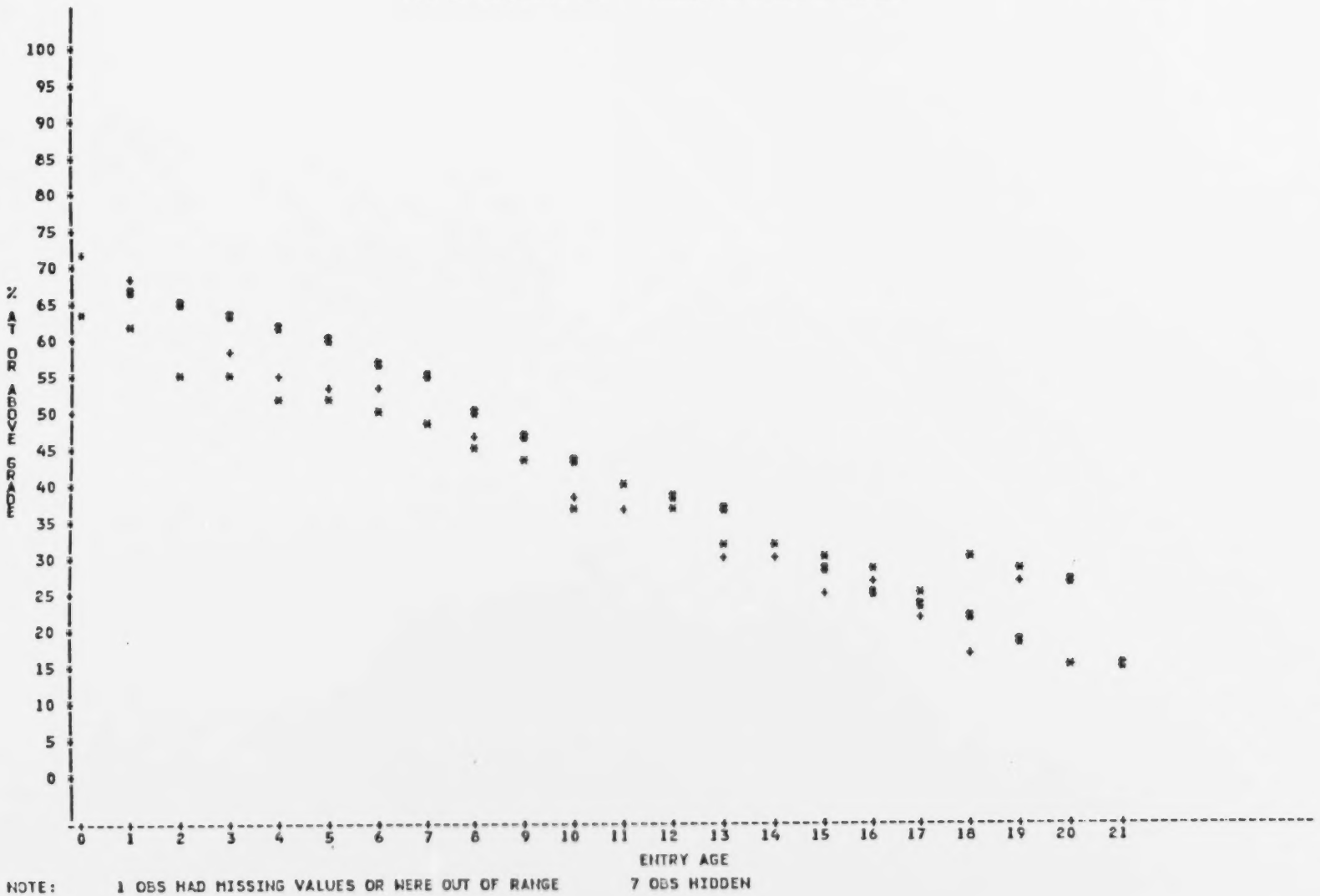
C.V.=Coefficient of variation=(root mse / mean) X 100

PR > F=Probability of obtaining an F value as large as the obtained F by chance.

From SASTDGLM 16:24 Thu, May 1, 1986

Table B3

PLOT OF PCT2*ENTR AGE SYMBOL IS VALUE OF SYMBOL



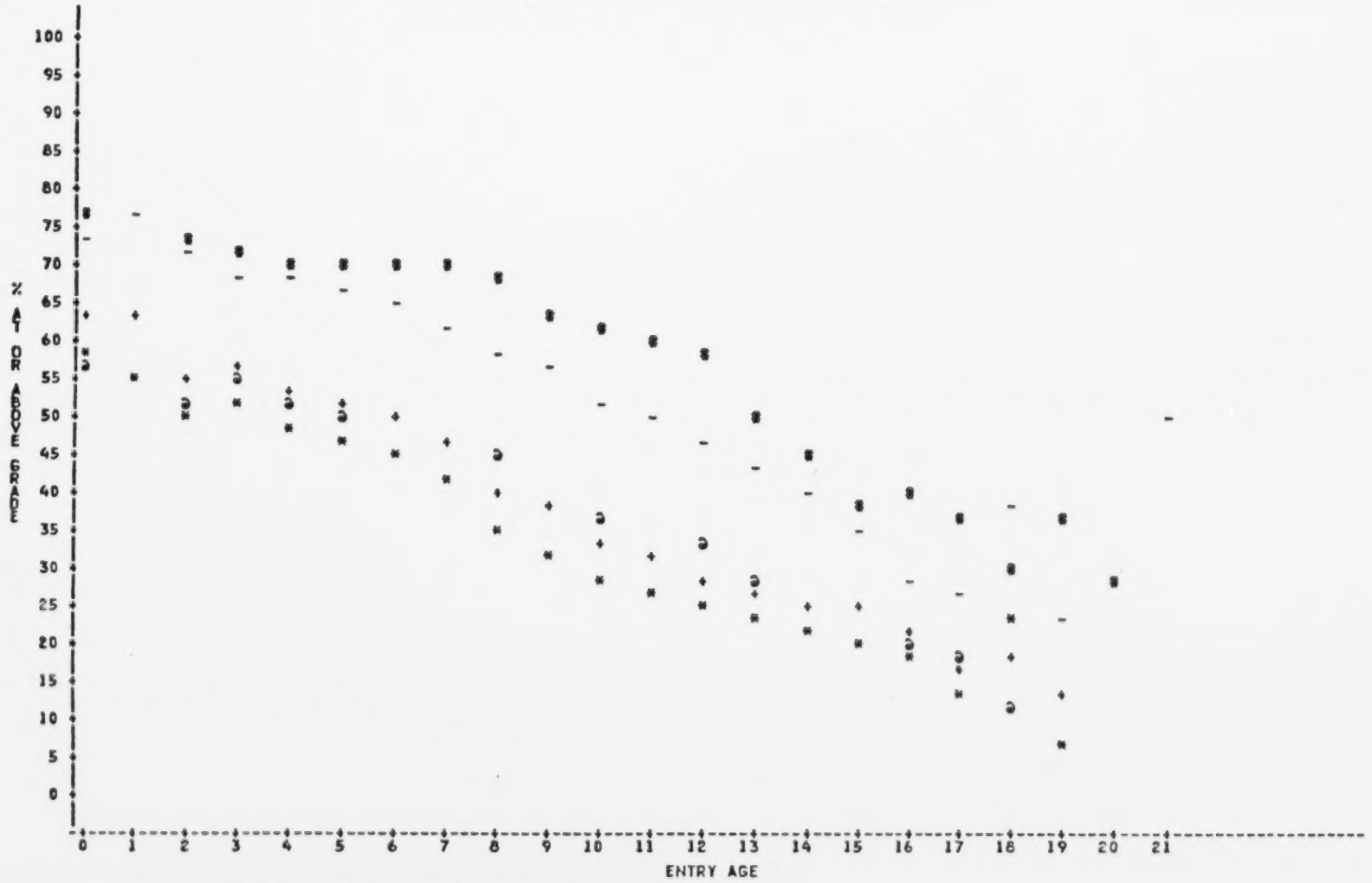
NOTE: 1 OBS HAD MISSING VALUES OR WERE OUT OF RANGE 7 OBS HIDDEN

PLOT B1

PERCENT AT OR ABOVE GRADE LEVEL BY ENTRY AGE
CTBS READING = * CTBS MATH = - CTBS LANGUAGE = +
CAT READING = 5 CAT MATH = 8

SASCGFR1

PLOT OF PCT2*ENTR AGE SYMBOL IS VALUE OF SYMBOL

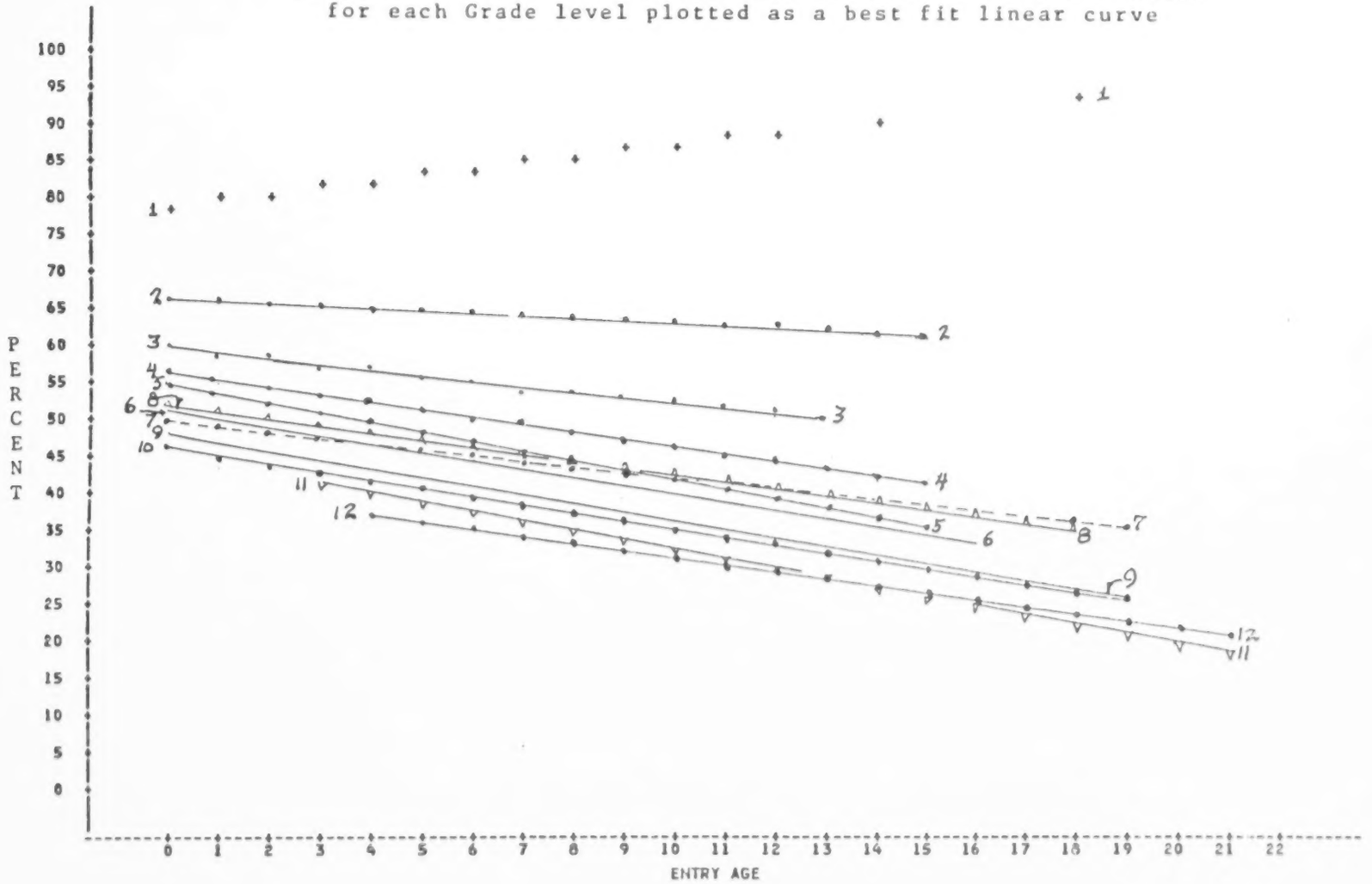


NOTE: 8 OBS HAD MISSING VALUES OR WERE OUT OF RANGE 9 OBS HIDDEN

PLOT B2

GRADE=1-12

Percent of (CTBS or CAT) test scores at or above Grade level
for each Grade level plotted as a best fit linear curve



PLOT B3

GLM REGRESSION OF MEAN ENTRY AGE ON GRADE SASCGM11 11:35 MONDAY, MAY 5, 1986 2

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: M ENTRY AGE
WEIGHT: N CASES

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	1082982.46119453	1082982.46119453	2446.48	0.0001	0.995929	237.7358
ERROR	10	4426.69736856	442.66973686		ROOT MSE		M ENTRY MEAN
CORRECTED TOTAL	11	1087409.15856308			21.03971808		8.85004241

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	1082982.46119453	2446.48	0.0001	1	1082982.46119453	2446.48	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	4.75091122	64.70	0.0001	0.07342464
GRADE	0.62601674	49.46	0.0001	0.01265654

TABLE C1

GLM REGRESSION OF MEAN GRADE ON ENTRY AGE SASCGM12 15:03 MONDAY, MAY 5, 1986 3

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: MGRADE
WEIGHT: N CASES

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	798097.29168801	798097.29168801	463.54	0.0001	0.958638	590.5918
ERROR	20	34434.91947348	1721.74597367		ROOT MSE		MGRADE MEAN
CORRECTED TOTAL	21	832532.21116148			41.49392695		7.02582204

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	798097.29168801	463.54	0.0001	1	798097.29168801	463.54	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	1.41941024	7.72	0.0001	0.18376187
ENTR AGE	0.46133920	21.53	0.0001	0.02142777

TABLE C2

MEAN OF ENTRY AGE BY GRADE SASCGMN1 11:35 MONDAY, MAY 5, 1986 1

VARIABLE	LABEL	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
					GRADE=1					
ENTR AGE	ENTRY AGE	34337	5.53161313	1.81573336	1.00000000	19.00000000	0.00979875	189939.0000	3.29688764	32.825
					GRADE=2					
ENTR AGE	ENTRY AGE	41886	6.05756100	1.94421558	1.00000000	16.00000000	0.00949970	253727.0000	3.77997423	32.096
					GRADE=3					
ENTR AGE	ENTRY AGE	42978	6.62673926	2.13587887	1.00000000	14.00000000	0.01030276	284804.0000	4.56197853	32.231
					GRADE=4					
ENTR AGE	ENTRY AGE	43891	7.17026270	2.36868683	1.00000000	16.00000000	0.01130628	314710.0000	5.61067730	33.035
					GRADE=5					
ENTR AGE	ENTRY AGE	42317	7.79835527	2.58715353	1.00000000	20.00000000	0.01257664	330003.0000	6.69336340	33.176
					GRADE=6					
ENTR AGE	ENTRY AGE	37674	8.40983171	2.85668529	1.00000000	17.00000000	0.01471775	316832.0000	8.16065082	33.968
					GRADE=7					
ENTR AGE	ENTRY AGE	34005	9.06016762	3.11246142	1.00000000	20.00000000	0.01687845	308091.0000	9.68741610	34.353
					GRADE=8					
ENTR AGE	ENTRY AGE	27198	9.64225311	3.35895954	1.00000000	19.00000000	0.02036743	262250.0000	11.28260921	34.836
					GRADE=9					
ENTR AGE	ENTRY AGE	18663	10.61072711	3.80032187	1.00000000	20.00000000	0.02781824	198028.0000	14.44244628	35.816
					GRADE=10					
ENTR AGE	ENTRY AGE	14295	11.24665967	3.90858782	1.00000000	20.00000000	0.03269097	160771.0000	15.27705877	34.753
					GRADE=11					
ENTR AGE	ENTRY AGE	9666	11.78957169	4.04631731	1.00000000	22.00000000	0.04115632	113958.0000	16.37268379	34.321
					GRADE=12					
ENTR AGE	ENTRY AGE	4101	12.25676664	4.29232511	1.00000000	22.00000000	0.06702668	50265.00000	18.42405481	35.020

TABLE C3

MEAN OF GRADE BY ENTRY AGE WEIGHTED SASCGMN2 15:03 MONDAY, MAY 5, 1986 1

VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
ENTRY AGE=1									
GRADE	3091	2.60465869	1.79669716	1.00000000	12.00000000	0.03231658	8051.0000000	3.22812068	68.980
ENTRY AGE=2									
GRADE	5920	2.70185811	1.67572028	1.00000000	10.00000000	0.02177914	15995.000000	2.80803845	62.021
ENTRY AGE=3									
GRADE	10265	3.33745738	2.03048300	1.00000000	11.00000000	0.02004102	34259.000000	4.12286123	60.839
ENTRY AGE=4									
GRADE	19105	3.59790631	2.17371514	1.00000000	11.00000000	0.01572639	68738.000000	4.72503749	60.416
ENTRY AGE=5									
GRADE	29403	3.91687923	2.39429159	1.00000000	12.00000000	0.01396308	115168.00000	5.73263223	61.128
ENTRY AGE=6									
GRADE	72114	4.23854846	2.55026477	1.00000000	12.00000000	0.00949676	305226.00000	6.50385038	60.254
ENTRY AGE=7									
GRADE	49372	4.51668962	2.71734339	1.00000000	12.00000000	0.01222937	222998.00000	7.38395511	60.162
ENTRY AGE=8									
GRADE	35287	4.67262731	2.57844227	1.00000000	12.00000000	0.01372619	164883.00000	6.64836453	55.182
ENTRY AGE=9									
GRADE	29608	5.17089976	2.42968636	1.00000000	12.00000000	0.01412036	153100.00000	5.90337580	46.988
ENTRY AGE=10									
GRADE	23839	5.76001510	2.24204503	1.00000000	12.00000000	0.01452113	137313.00000	5.02676594	38.924
ENTRY AGE=11									
GRADE	20567	6.36077211	2.08889143	1.00000000	12.00000000	0.01456567	130822.00000	4.36346742	32.840
ENTRY AGE=12									
GRADE	15856	6.95547427	1.92398207	1.00000000	12.00000000	0.01527933	110286.00000	3.70170702	27.661
ENTRY AGE=13									
GRADE	11997	7.61348671	1.79786139	1.00000000	12.00000000	0.01641421	91339.000000	3.23230559	23.614
ENTRY AGE=14									
GRADE	9151	8.29953011	1.67096674	2.00000000	12.00000000	0.01746761	75949.000000	2.79212983	20.133
ENTRY AGE=15									
GRADE	6310	8.98462758	1.51578104	1.00000000	12.00000000	0.01908191	56693.000000	2.29759215	16.871
ENTRY AGE=16									
GRADE	4497	9.62263731	1.27568064	2.00000000	12.00000000	0.01902307	43273.000000	1.62736110	13.257
ENTRY AGE=17									
GRADE	2612	10.17343032	1.13986233	6.00000000	12.00000000	0.02230313	26573.000000	1.29926612	11.204
ENTRY AGE=18									
GRADE	1347	10.60430586	1.04579219	7.00000000	12.00000000	0.02849454	14284.000000	1.09368130	9.862
ENTRY AGE=19									
GRADE	499	10.94188377	1.20350445	1.00000000	12.00000000	0.05387626	5460.0000000	1.44842295	10.999
ENTRY AGE=20									
GRADE	144	10.98611111	1.32214125	5.00000000	12.00000000	0.11017844	1582.0000000	1.74805750	12.035
ENTRY AGE=21									
GRADE	20	11.80000000	0.41039134	11.00000000	12.00000000	0.09176629	236.00000000	0.16842105	3.478
ENTRY AGE=22									
GRADE	7	11.71428571	0.48795004	11.00000000	12.00000000	0.18442778	82.00000000	0.23809524	4.165

TABLE C4

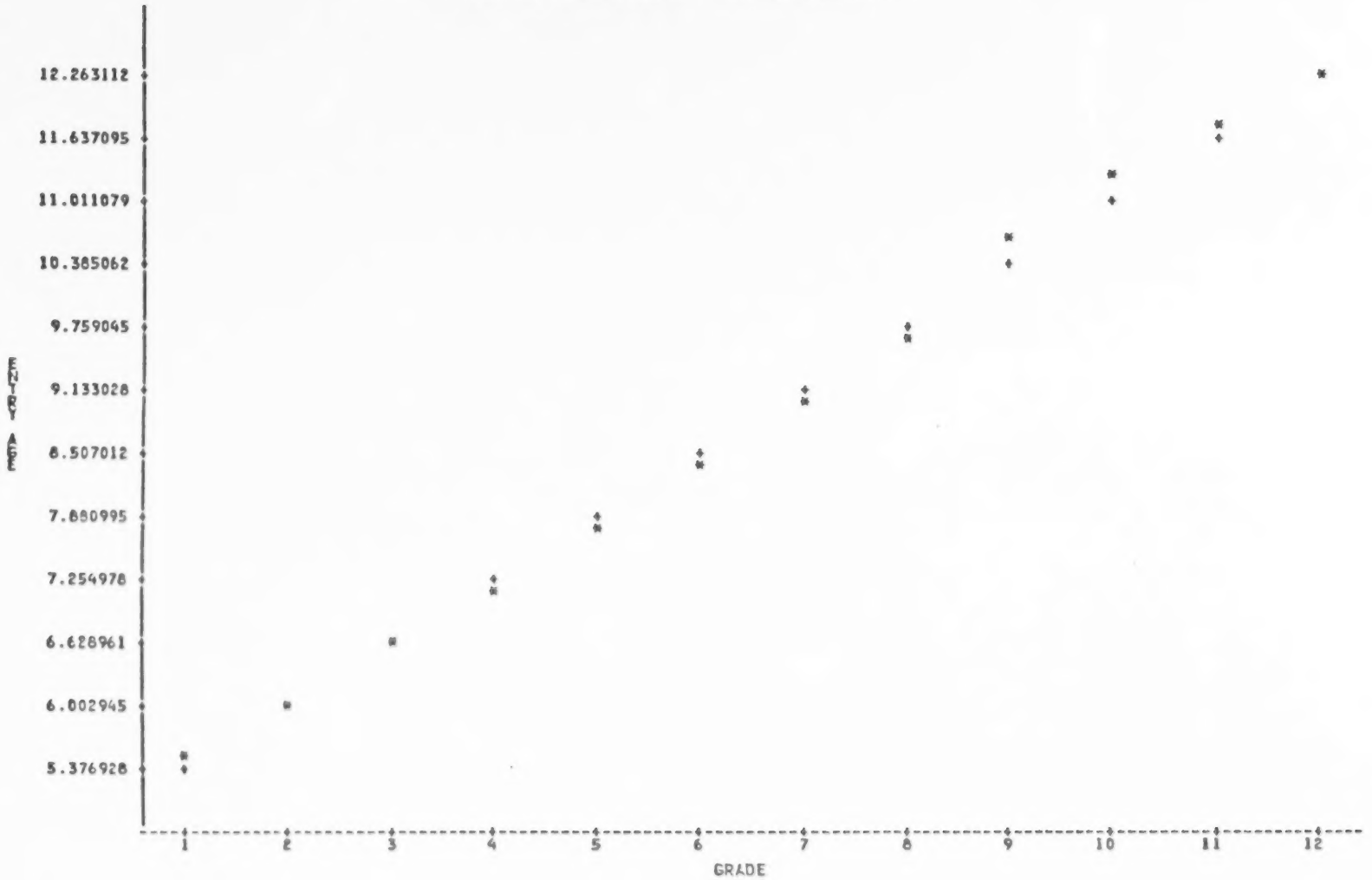
II

PLOT OF MEAN ENTRY AGE BY GRADE AND PREDICTED ENTRY AGE BY GRADE

11:35 MONDAY, MAY 5, 1986 3

SASCGM01

PLOT OF M_ENTRY*GRADE SYMBOL USED IS *
PLOT OF PREDICT*GRADE SYMBOL USED IS +



PLOT C1

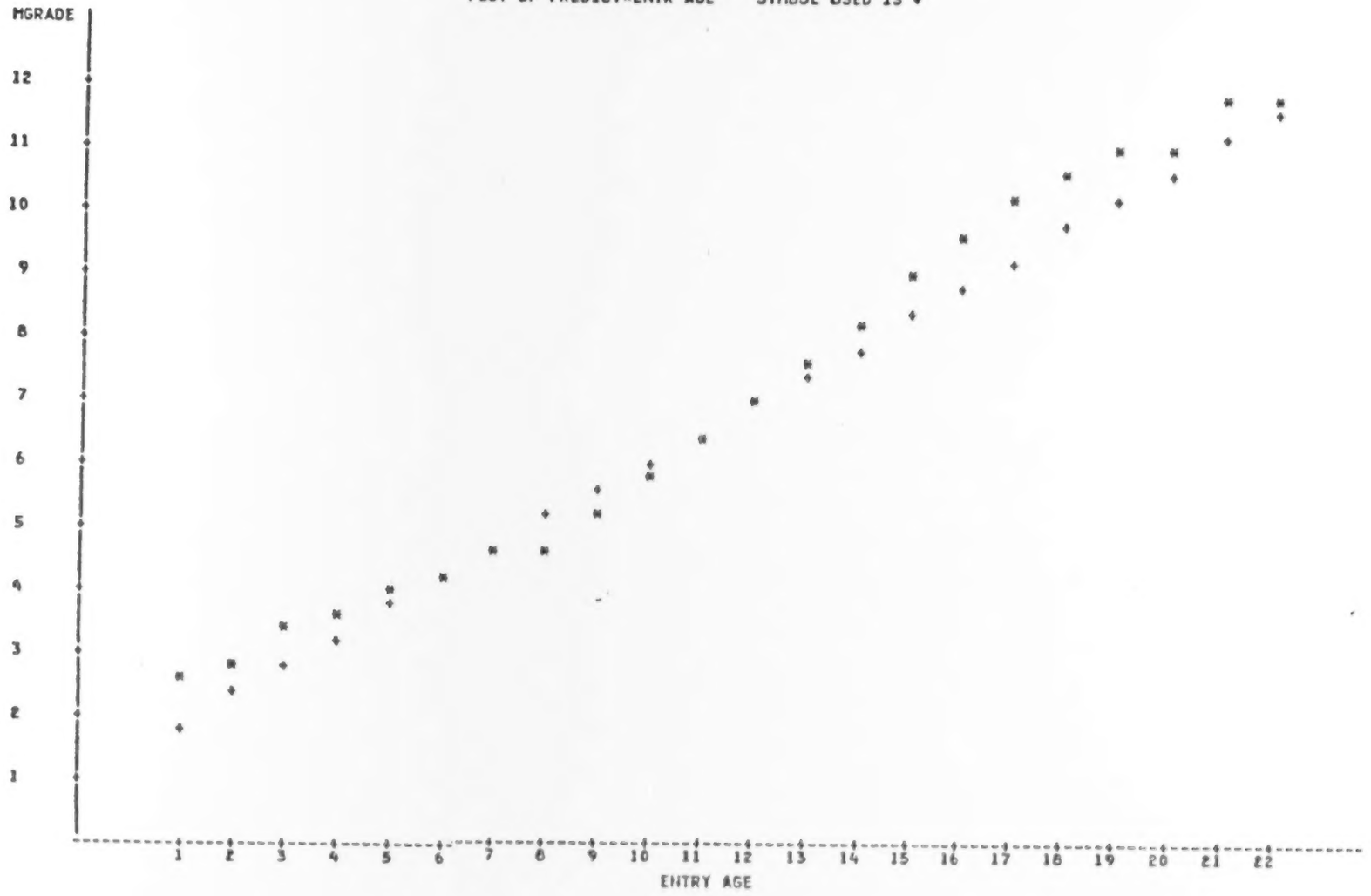
12

PLOT OF GRADE MEAN BY ENTRY AGE AND
PREDICTED GRADE BY ENTRY AGE WEIGHTED

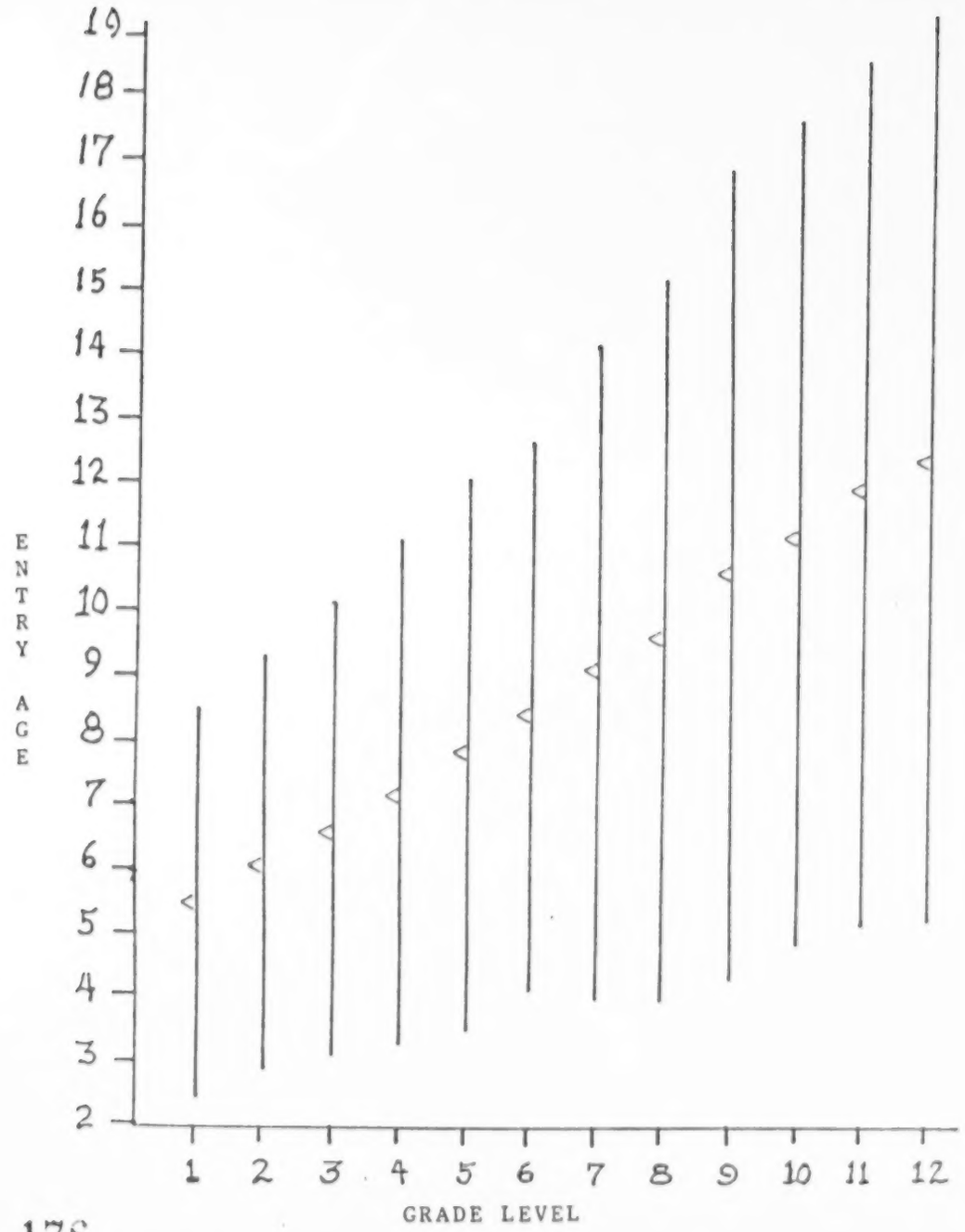
15:03 MONDAY, MAY 5, 1986 4

SASCGMN2

PLOT OF MGRADE*ENTR AGE SYMBOL USED IS *
PLOT OF PREDICT*ENTR AGE SYMBOL USED IS +



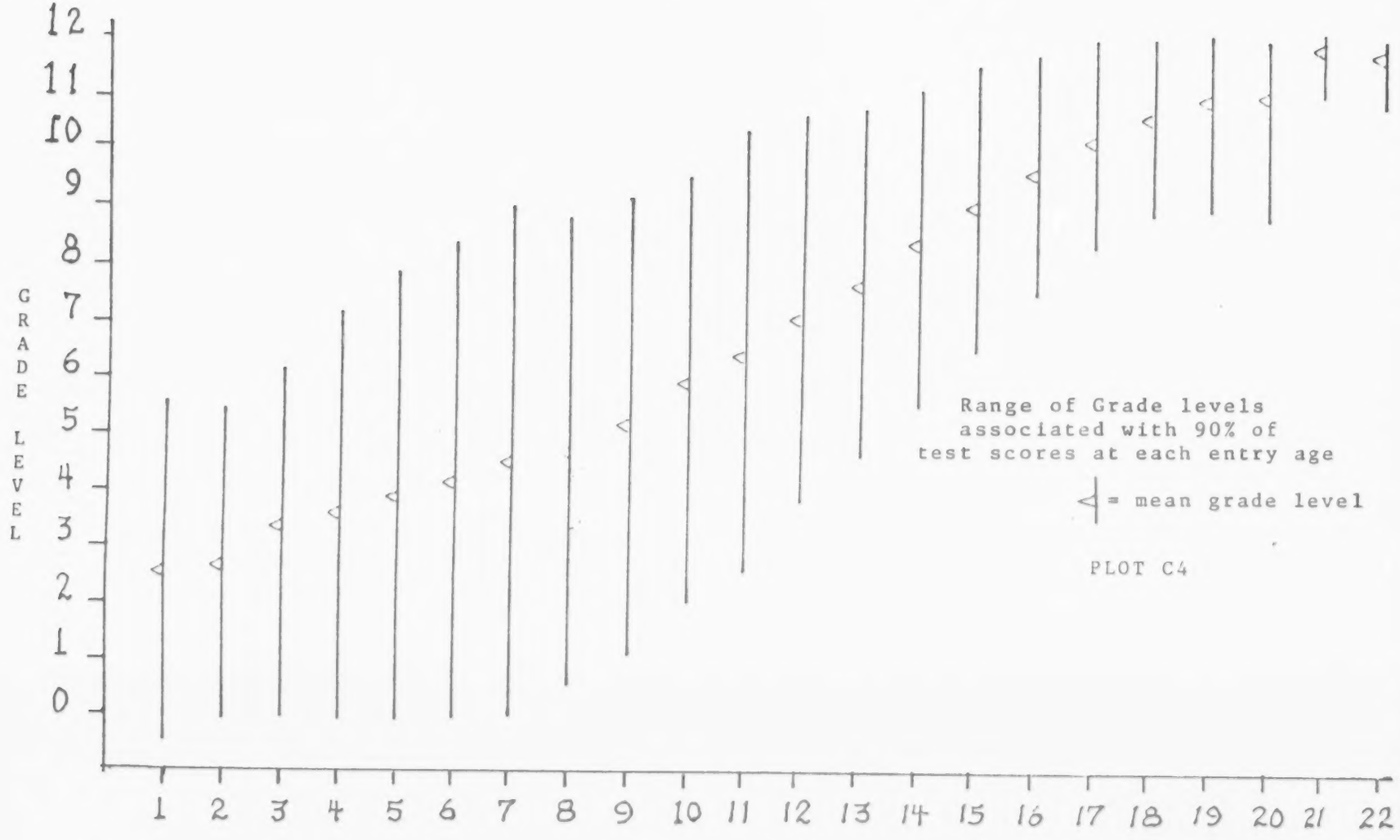
PLOT C2

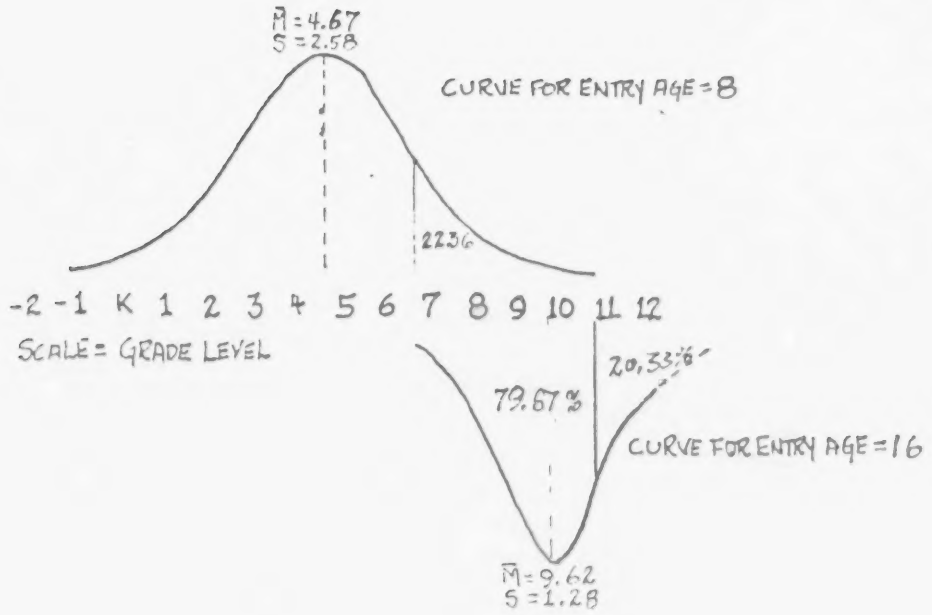


Range of Entry ages
associated with 90% of
test scores at each grade level

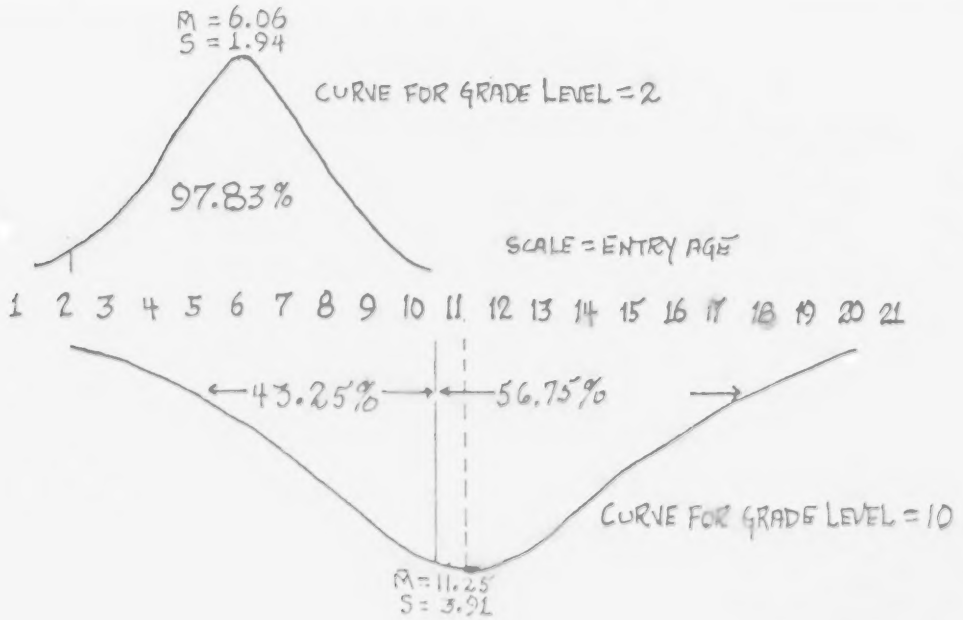
= mean entry age

PLOT C3

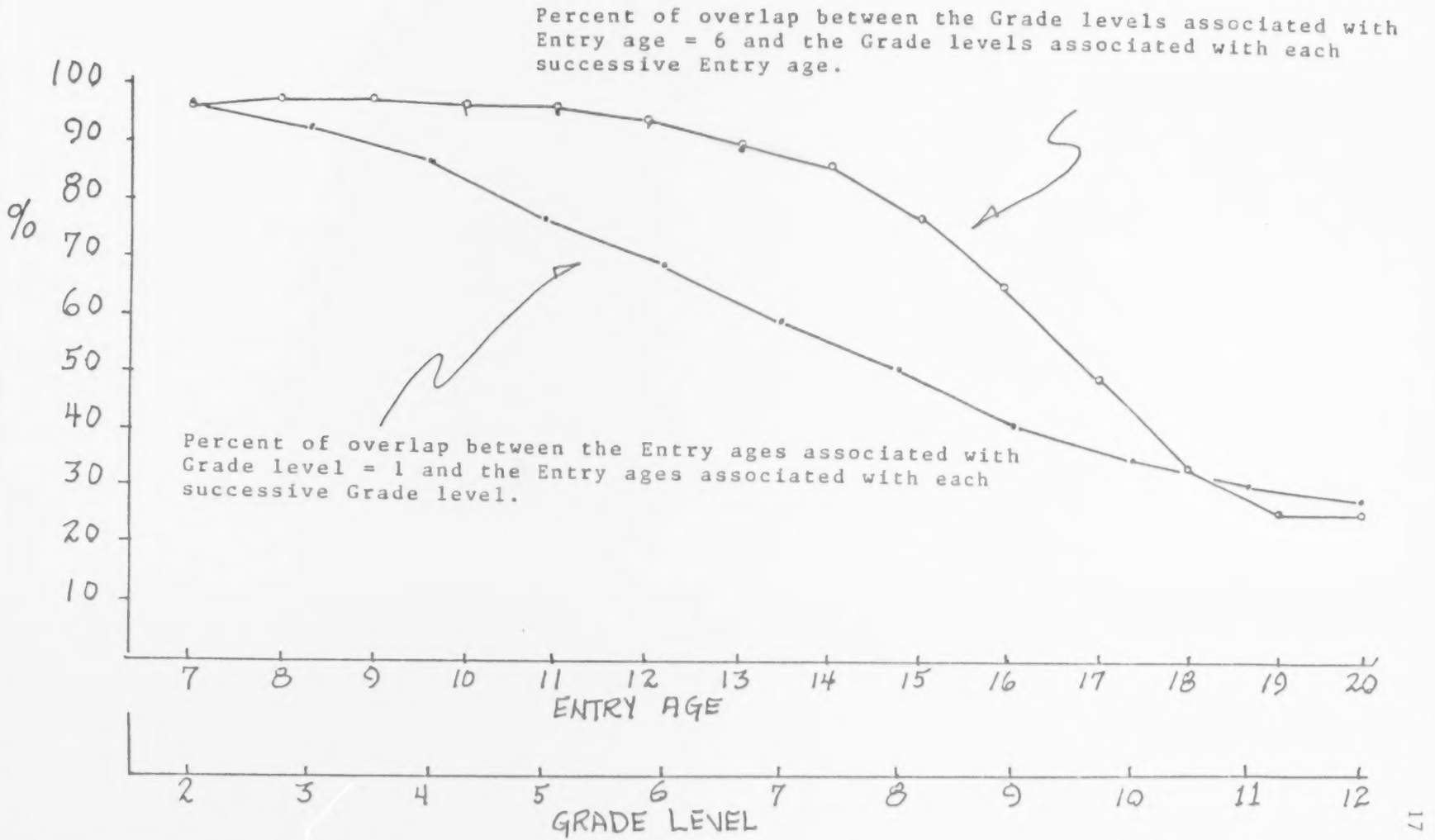

GRADE LEVEL
12
11
10
9
8
7
6
5
4
3
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
ENTRY AGE
Range of Grade levels
associated with 90% of
test scores at each entry age
= mean grade level
PLOT C4



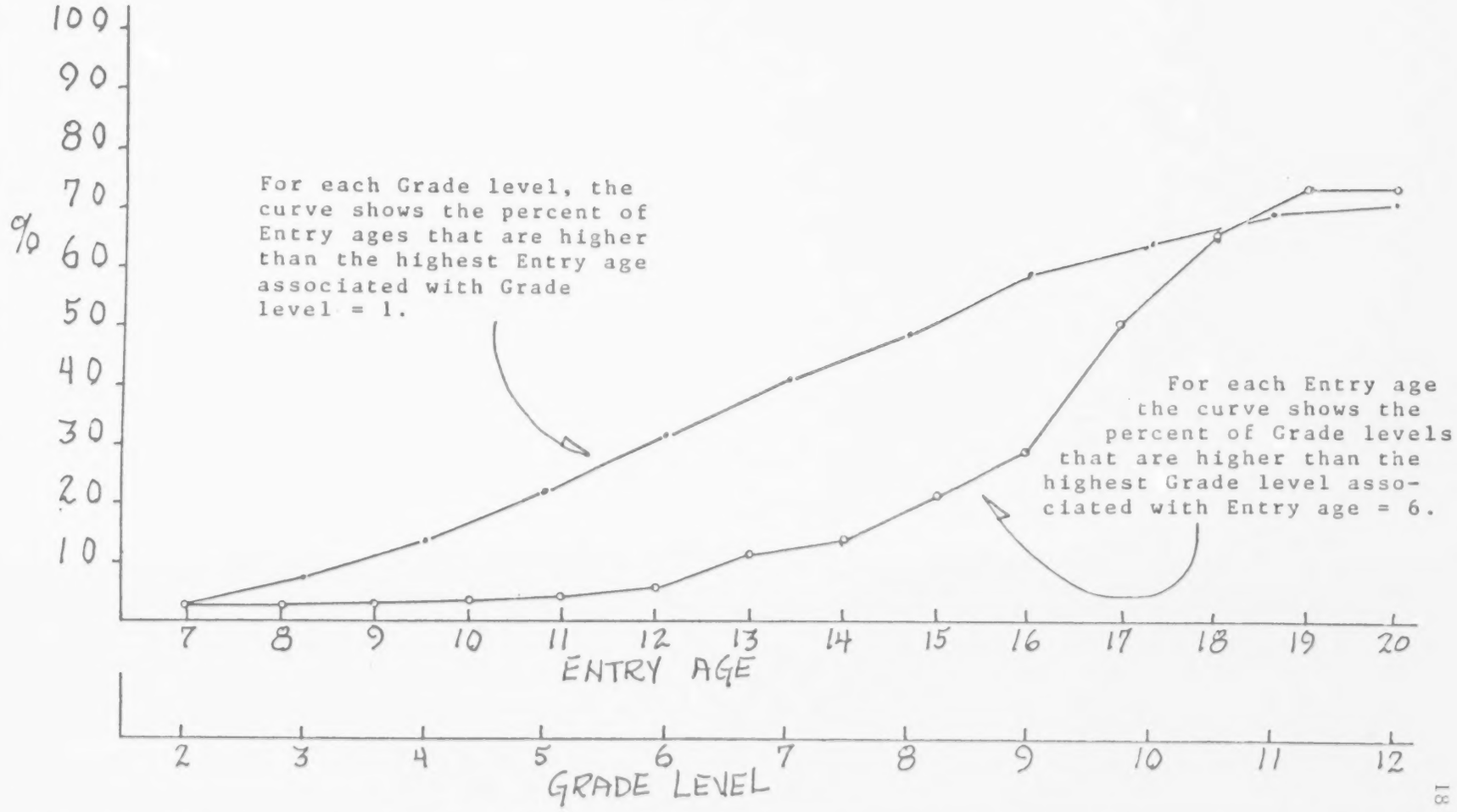
PLOT C5-1



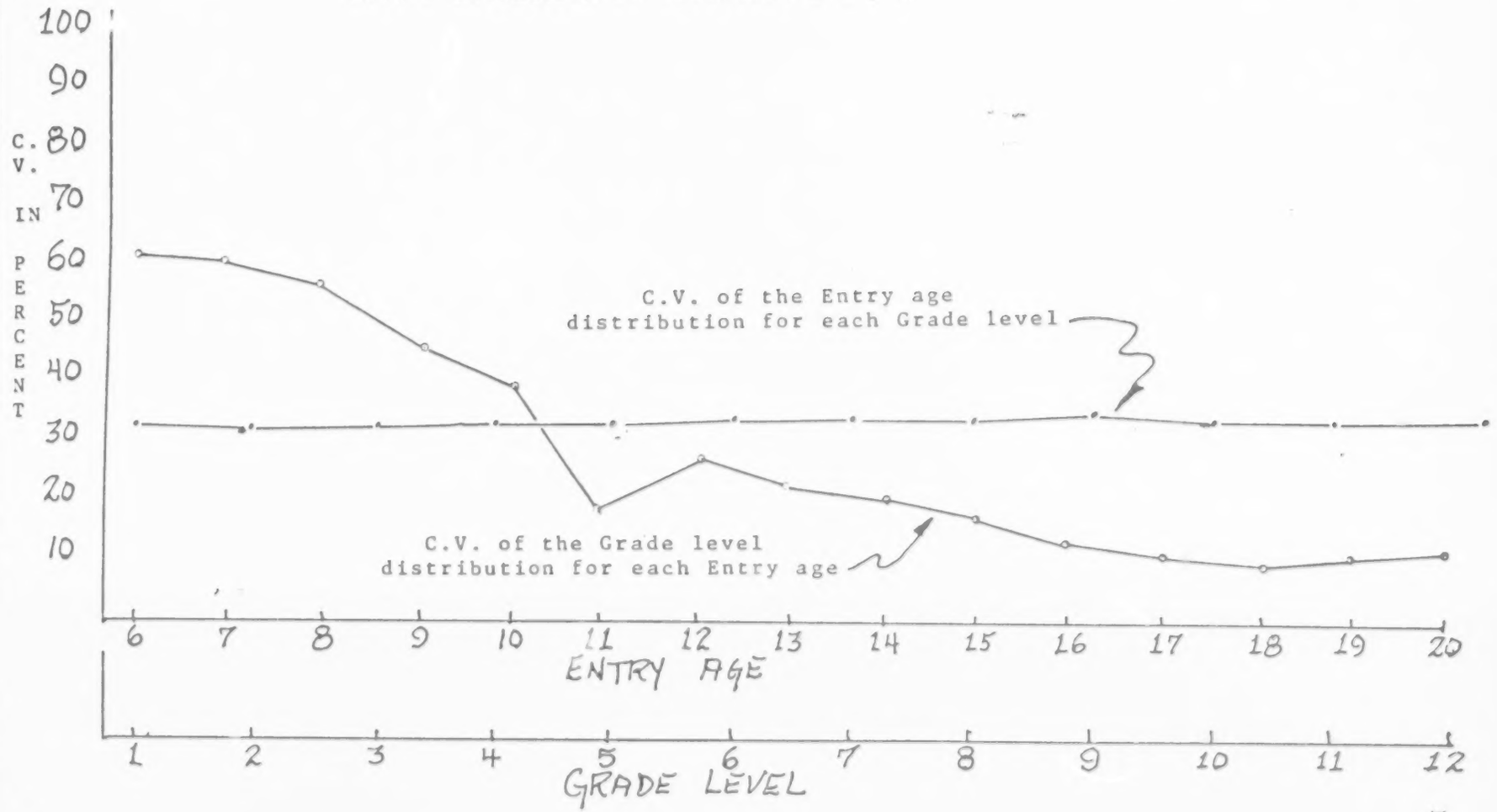
PLOT C5-2


Percent of overlap between the Grade levels associated with Entry age = 6 and the Grade levels associated with each successive Entry age.
Percent of overlap between the Entry ages associated with Grade level = 1 and the Entry ages associated with each successive Grade level.
%
100
90
80
70
60
50
40
30
20
10
7 8 9 10 11 12 13 14 15 16 17 18 19 20
ENTRY AGE
2 3 4 5 6 7 8 9 10 11 12
GRADE LEVEL

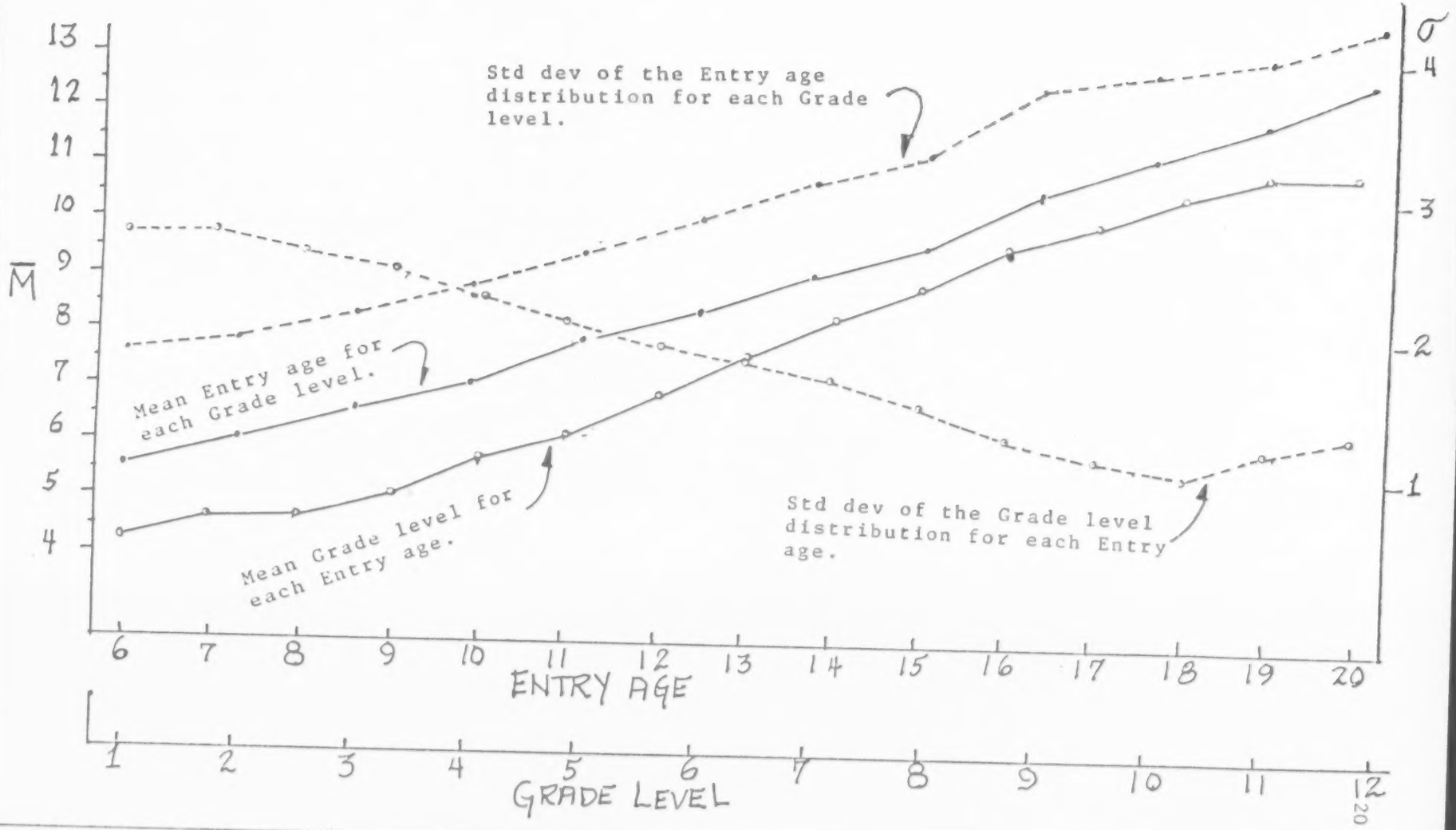
PLOT C6


100
90
80
70
60
50
40
30
20
10
%
For each Grade level, the curve shows the percent of Entry ages that are higher than the highest Entry age associated with Grade level = 1.
For each Entry age the curve shows the percent of Grade levels that are higher than the highest Grade level associated with Entry age = 6.
7 8 9 10 11 12 13 14 15 16 17 18 19 20
ENTRY AGE
2 3 4 5 6 7 8 9 10 11 12
GRADE LEVEL

C.V. = Coeficient of variation = σ / M



PLOT C8


13
12
11
10
9
8
7
6
5
4
$\bar{M}$
σ
4
3
2
1
Std dev of the Entry age distribution for each Grade level.
Mean Entry age for each Grade level.
Mean Grade level for each Entry age.
Std dev of the Grade level distribution for each Entry age.
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
ENTRY AGE
1 2 3 4 5 6 7 8 9 10 11 12
GRADE LEVEL
120

PLOT C9

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
STATUS	3	1 2 3
TEST TYP	5	101 102 103 301 302
GRADE	12	1 2 3 4 5 6 7 8 9 10 11 12

NUMBER OF OBSERVATIONS IN DATA SET = 2859

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	179	10214.89786659	57.06646853	133.23	0.0001	0.909914	147.8503
ERROR	2361	1011.32214638	0.42834483		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	2540	11226.22001298			0.65448058		0.44266424

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
STATUS	2	118.10593623	137.86	0.0001	2	16.55075903	19.32	0.0001
TEST TYP	4	3451.61651012	2014.51	0.0001	4	1414.65441183	825.65	0.0001
GRADE	11	6179.82622161	1311.57	0.0001	11	3912.37734591	830.34	0.0001
STATUS*TEST TYP	8	52.75259511	3.72	0.0002	8	7.88099801	2.30	0.0167
STATUS*GRADE	22	52.99640403	5.50	0.0001	22	26.93211695	2.86	0.0001
TEST TYP*GRADE	44	357.58684425	18.97	0.0001	44	262.42672131	13.92	0.0001
STATUS*TEST T*GRADE	88	62.01135524	1.65	0.0002	88	62.01135524	1.65	0.0002

TABLE D1

MEANS OF MAIN EFFECT VARIABLES USED IN
ANALYSIS OF VARIANCE

VARIABLE	MEAN (in percent at or above grade level)
STATUS	
1	45.448
2	43.856
3	43.421
TEST TYPE	
CTBS Reading	32.987
CTBS Arithmetic	51.927
CTBS Language	39.482
CAT Reading	38.339
CAT Arithmetic	58.005
Grade Level	
1	80.115
2	62.884
3	53.613
4	47.670
5	45.855
6	41.905
7	42.153
8	42.839
9	35.630
10	34.363
11	32.592
12	30.910

TABLE D2

STATUS=1

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	7306318.73812776	7306318.73812776	36.60	0.0001	0.785392	964.8103
ERROR	10	1996446.21218311	199644.62121831		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	9302764.95031086			446.81609328		46.31128856

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	7306318.73812776	36.60	0.0001	1	7306318.73812776	36.60	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	67.97798461	18.82	0.0001	3.61197469
GRADE	-3.56229509	-6.05	0.0001	0.58885639

STATUS=2

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	8579797.48531141	8579797.48531141	29.35	0.0003	0.745864	1211.5476
ERROR	10	2923371.17780253	292337.11778025		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	11503168.66311390			540.68208568		44.62739260

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	8579797.48531141	29.35	0.0003	1	8579797.48531141	29.35	0.0003

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	71.00506909	15.18	0.0001	4.67688302
GRADE	-4.30291423	-5.42	0.0003	0.79426633

STATUS=3

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	34013057.87299910	34013057.87299910	45.59	0.0001	0.820123	1820.7441
ERROR	10	7460083.12613373	746008.31261337		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	41473140.99913280			863.71772739		47.43762216

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	34013057.87299910	45.59	0.0001	1	34013057.87299910	45.59	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	74.71337451	19.96	0.0001	3.74398227
GRADE	-4.45462216	-6.75	0.0001	0.65972033

23

TABLE E1

TEST TYP=101

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	15130114.29226820	15130114.29226820	25.62	0.0005	0.719231	2147.1463
ERROR	10	5906388.06253371	590638.80625337		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	21036502.35480200			768.53028974		35.79310348

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	15130114.29226820	25.62	0.0005	1	15130114.29226820	25.62	0.0005

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	64.58141346	12.00	0.0001	5.38177364
GRADE	-4.83111318	-5.06	0.0005	0.95452454

TEST TYP=102

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	11571348.11321360	11571348.11321360	174.87	0.0001	0.945909	471.7088
ERROR	10	661700.06934033	66170.00693403		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	12233048.18255400			257.23531432		54.53265544

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	11571348.11321360	174.87	0.0001	1	11571348.11321360	174.87	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	82.71983515	44.67	0.0001	1.85198874
GRADE	-4.42665729	-13.22	0.0001	0.33474532

TEST TYP=103

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	7829728.57104615	7829728.57104615	34.82	0.0002	0.776867	1130.6187
ERROR	10	2248864.94780338	224886.49478034		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	10078593.51884950			474.22198893		41.94358108

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	7829728.57104615	34.82	0.0002	1	7829728.57104615	34.82	0.0002

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	66.22476315	16.79	0.0001	3.94367386
GRADE	-4.06034506	-5.90	0.0002	0.68813124

24

TABLE E2 (cont'd next page)

TEST TYP=301

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	8480259.06840320	8480259.06840320	15.00	0.0031	0.599992	1797.6775
ERROR	10	5653697.69906909	565369.76990691		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	14133956.76747230			751.91074597		41.82678646

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	8480259.06840320	15.00	0.0031	1	8480259.06840320	15.00	0.0031

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	65.84768743	10.38	0.0001	6.34357939
GRADE	-4.01486985	-3.87	0.0031	1.03665251

TEST TYP=302

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	8518586.08491138	8518586.08491138	130.56	0.0001	0.928858	414.0174
ERROR	10	652448.44885672	65244.84488567		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	9171034.53376810			255.43070467		61.69563827

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	8518586.08491138	130.56	0.0001	1	8518586.08491138	130.56	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	87.92952288	39.95	0.0001	2.20109006
GRADE	-4.10193269	-11.43	0.0001	0.35898630

TABLE E2 (cont'd)

Synopsis of data from the analysis of regression of student performance on Grade level in each Entry age

Ent age	Mean	Intcpt	Slope	R-Sq	C.V.	PR > F
01	56.02	77.71	-4.79	.73	252.73	.0034
02	49.16	78.60	-5.69	.85	336.48	.0001
03	56.27	77.32	-4.77	.76	477.61	.0005
04	50.45	77.70	-5.04	.77	700.59	.0002
05	49.70	75.83	-4.46	.74	1140.55	.0003
06	49.60	74.07	-4.16	.73	974.47	.0004
07	48.65	69.93	-3.64	.71	727.44	.0006
08	48.15	64.14	-2.97	.65	601.95	.0016
09	46.37	57.09	-2.07	.68	334.26	.0009
10	43.12	51.21	-1.41	.61	247.21	.0027
11	39.85	50.52	-1.57	.58	268.52	.0062
12	42.35	43.42	-0.68	.10	314.96	.3418***
13	33.26	47.75	-1.62	.29	424.77	.1050***
14	28.34	33.20	-0.16	.01	362.69	.8559***
15	23.20	35.59	-0.81	.08	407.79	.4927***
16	22.20	11.53	+1.42	.21	326.49	.3048***
17	23.95	8.99	+1.35	.13	327.67	.5588***
18	21.27	-.14	+2.12	.26	222.99	.3820***
19	25.79	43.69	-1.98	.24	115.88	.4004***

*** = Slope not significant at .05 level

Mean=Mean percent of test scores at or above grade level across all entry ages

R-Sq=R-Square=Mean square / corrected total sum of squares

C.V.=Coefficient of variation=(root mse / mean) X 100

PR > F=Probability of obtaining an F value as large as the obtained F by chance.

From SASTDG18 16:43 Mon, May 5, 1986

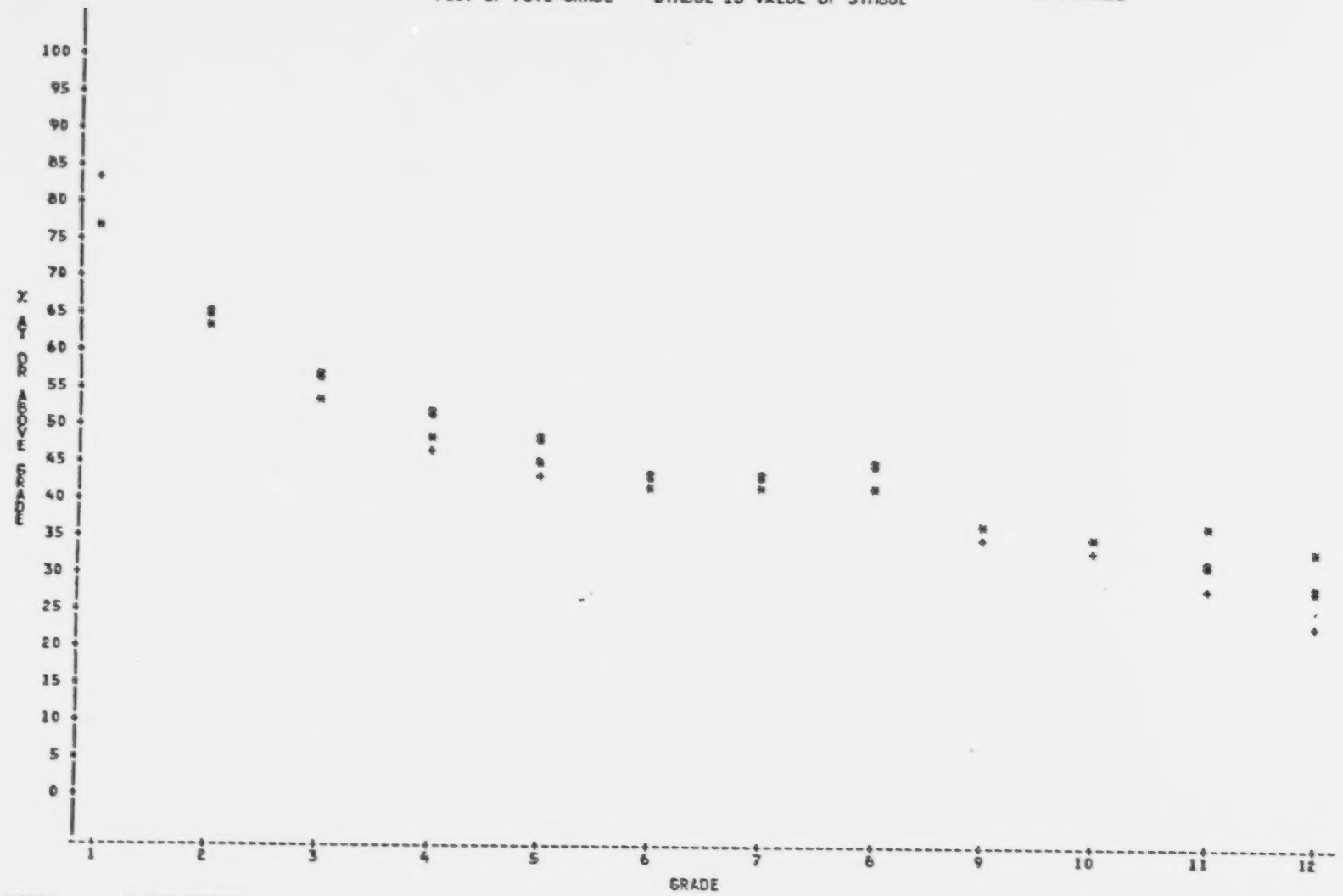
TABLE E3

PERCENT AT OR ABOVE GRADE LEVEL BY GRADE
* = STATUS 1 + = STATUS 2 8 = STATUS 3

7:55 MONDAY, APRIL 28, 1986 4

PLOT OF PCT2*GRADE SYMBOL IS VALUE OF SYMBOL

SASCGGL3



NOTE: 8 OBS HIDDEN

27

200

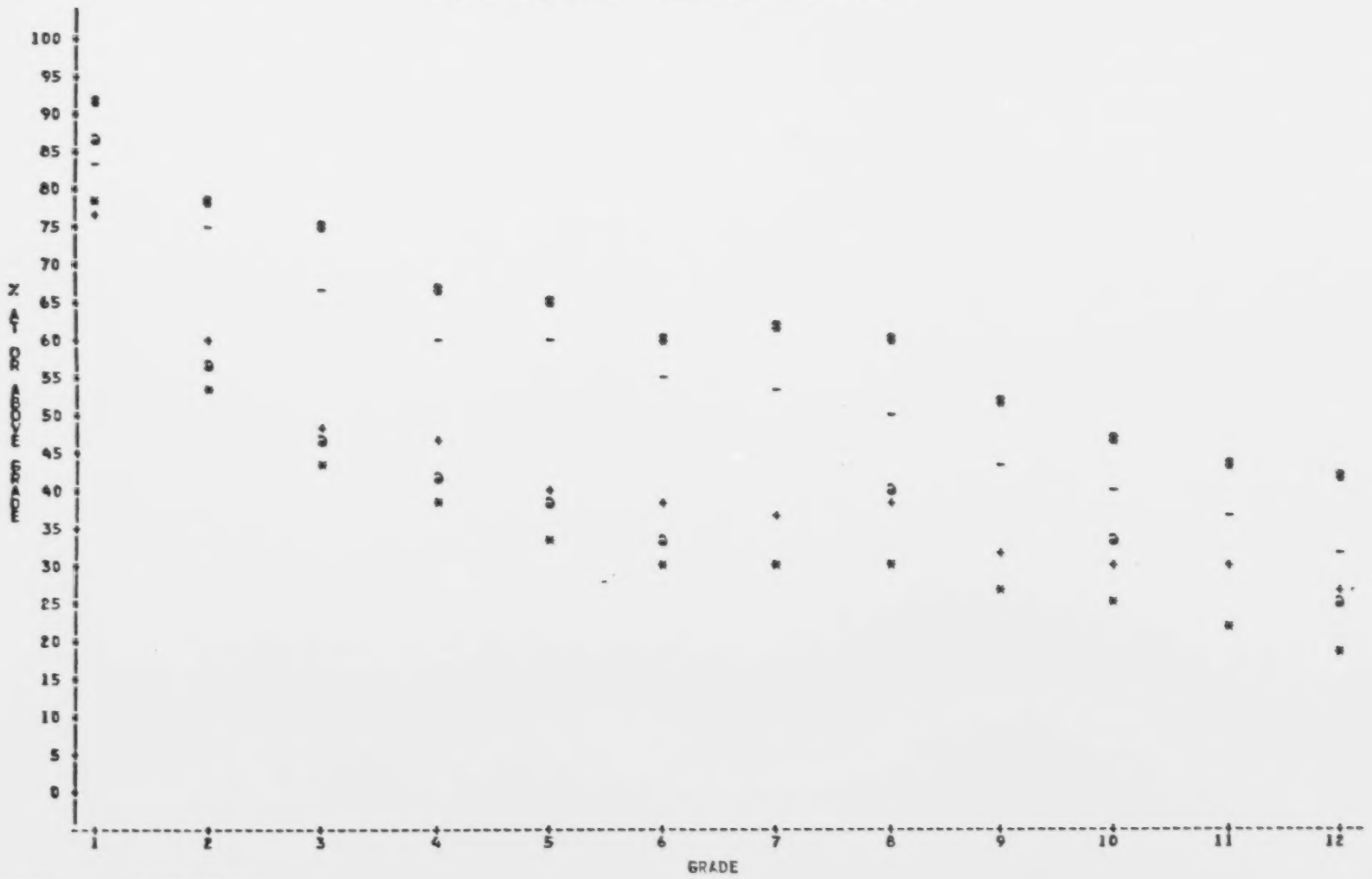
PLOT E1

199

CTBS READING = * CTBS MATH = - CTBS LANGUAGE = +
CAT READING = 9 CAT MATH = 8

SASCGFR2

PLOT OF PCT2*GRADE SYMBOL IS VALUE OF SYMBOL

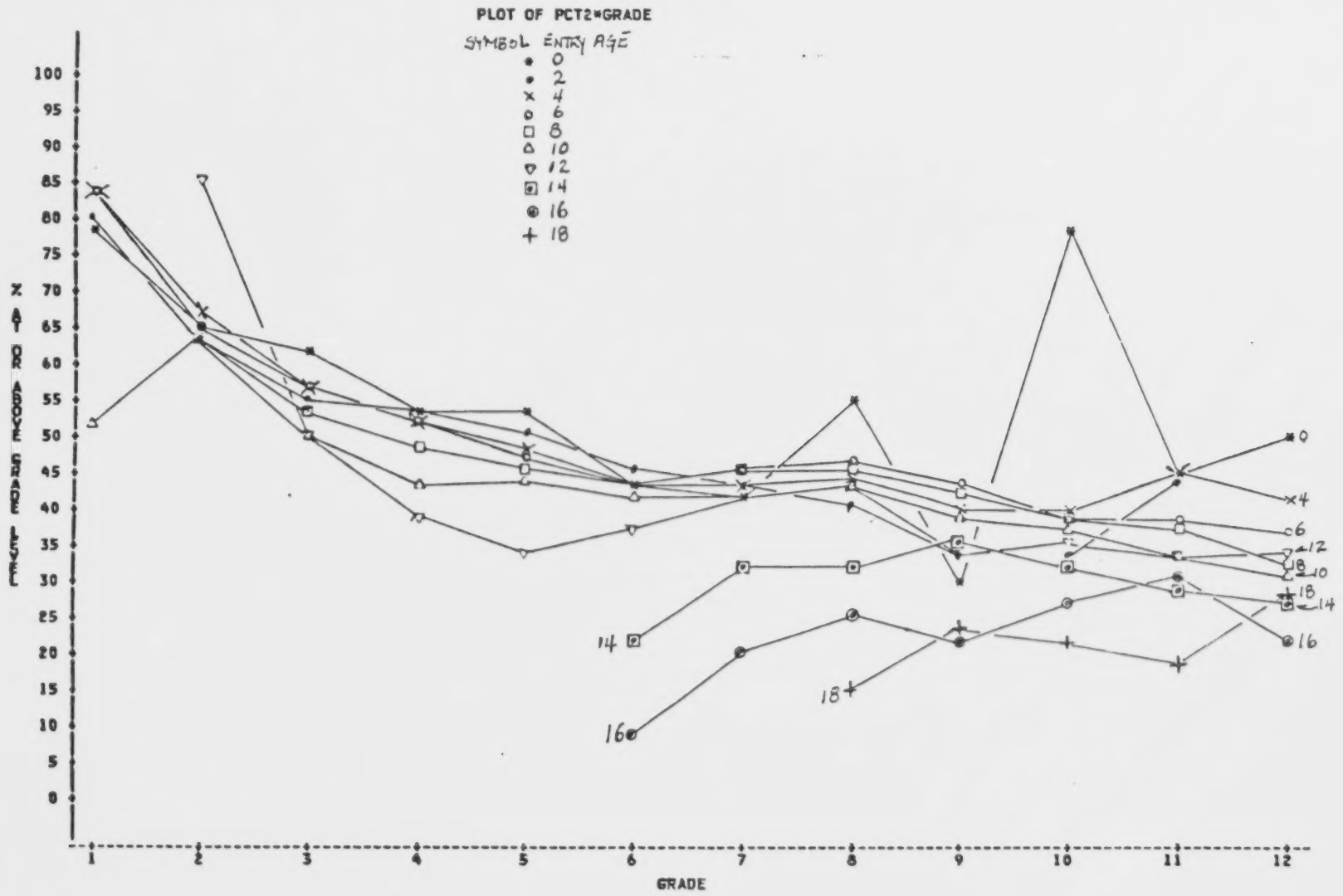


NOTE: 3 OBS HIDDEN

201

PLOT E2

202


PLOT OF PCT2*GRADE
SYMBOL ENTRY AGE
0
2
4
6
8
10
12
14
16
18
% AT OR ABOVE GRADE LEVEL
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
GRADE

PLOT E3

GLM ANOVA OF PCT2 = EXPGR GRADE TEST TYP STATUS SASEXGL8 11:27 FRIDAY, MAY 9, 1986 1

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
EXPGR	2	1 6
GRADE	12	1 2 3 4 5 6 7 8 9 10 11 12
TEST TYP	5	101 102 103 301 302
STATUS	3	1 2 3

NUMBER OF OBSERVATIONS IN DATA SET = 1115

NOTE: ALL DEPENDENT VARIABLES ARE CONSISTENT WITH RESPECT TO THE PRESENCE OR ABSENCE OF MISSING VALUES. HOWEVER, ONLY 770 OBSERVATIONS CAN BE USED IN THIS ANALYSIS.

TABLE F1

GLM ANOVA OF PCT2 = EXPGR GRADE TEST TYP STATUS SASEXGL8 11:27 FRIDAY, MAY 9, 1986 2

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	270	1877.88136304	6.95511616	31.41	0.0001	0.944437	107.2902
ERROR	499	110.47907162	0.22140095		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	769	1988.36043466			0.47053262		0.43856074

	SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
	EXPGR	1	111.23000962	502.39	0.0001	1	12.00553996	54.23	0.0001
	GRADE	11	1084.29391631	445.22	0.0001	11	469.20793105	192.66	0.0001
	TEST TYP	4	524.62496881	592.39	0.0001	4	94.72708568	106.99	0.0001
	STATUS	2	2.20293573	4.97	0.0073	2	1.90978953	4.31	0.0139
	EXPGR*GRADE	11	11.96481060	4.91	0.0001	11	6.83830570	2.81	0.0015
	EXPGR*TEST TYP	4	3.36005001	3.79	0.0047	4	2.32303403	2.62	0.0341
	EXPGR*STATUS	2	0.11535860	20.59	0.0001	2	7.78541013	17.56	0.0001
	GRADE*TEST TYP	44	73.78600676	7.57	0.0001	44	33.91680181	3.48	0.0001
	GRADE*STATUS	22	9.70416529	1.99	0.0049	22	8.76194480	1.80	0.0147
1	TEST TYP*STATUS	8	3.23678468	1.63	0.0695	8	2.19183275	1.24	0.2749
1	EXPGR*GRADE*TEST TY	43	7.31824632	0.77	0.8561	43	7.89203488	0.83	0.7729
	EXPGR*GRADE*STATUS	22	10.07016575	2.07	0.0032	22	10.59321835	2.17	0.0017
	EXPGR*TEST T*STATUS	8	4.64663015	2.62	0.0090	8	4.85300805	2.74	0.0058
1	GRADE*TEST T*STATUS	88	22.32531440	1.15	0.1886	88	22.32531440	1.15	0.1886

1/ NOT SIGNIFICANT AT .05 LEVEL

30

205

TABLE F2

206

MEANS OF MAIN EFFECT VARIABLES
from GLM analysis of variance

VARIABLE	MEAN (in percent at or above grade level)
EXP GROUP	
6	49.02
1	42.28
STATUS	
1	46.19
2	43.32
3	42.22
TEST TYPE	
CTBS Reading	40.29
CTBS Arithmetic	49.07
CTBS Language	40.29
CAT Reading	37.87
CAT Arithmetic	54.34
Grade Level	
1	77.29
2	55.24
3	47.70
4	44.57
5	39.66
6	39.92
7	41.91
8	39.97
9	33.12
10	34.33
11	32.30

From SASEXGL8 11:27 Fri, May 9, 1986

TABLE F3

Data used to analyze the regression of Percent of scores at or above grade on Experimental Group 1 and Experimental Group 6

SASEXGL5
8:50 Fri,
May 9, 86

OBSERVATION NUMBER	EXPERIMENTAL GROUP	GRADE LEVEL	NUMBER OF CASES IN OBSERVATION
1	1	1	1241
2	1	2	1775
3	1	3	1708
4	1	4	1543
5	1	5	1410
6	1	6	1208
7	1	7	1058
8	1	8	824
9	1	9	1039
10	1	10	480
11	1	11	294
12	1	12	147
13	6	1	6567
14	6	2	7167
15	6	3	7232
16	6	4	6679
17	6	5	5666
18	6	6	4439
19	6	7	3783
20	6	8	3055
21	6	9	1893
22	6	10	1484
23	6	11	961
24	6	12	446

TABLE F4

MEAN PERCENT OF SCORES AT OR ABOVE GRADE LEVEL
FOR EACH OF TWO EXPERIMENTAL GROUPS
BY GRADE LEVEL

	GRADE LEVEL											
	1	2	3	4	5	6	7	8	9	10	11	12
EXPGR6	83.55	64.35	56.04	52.48	47.23	44.13	44.30	45.96	43.32	39.15	38.29	36.32
EXPGR1	81.47	51.94	46.96	40.44	37.87	35.93	37.81	33.01	28.87	25.00	18.37	21.09

TABLE F5

EXPGR=6

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	629.53095964	629.53095964	26.95	0.0004	0.729391	974.4675
ERROR	10	233.55973880	23.35597388		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	863.09069843			4.83280187		0.49594286

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	629.53095964	26.95	0.0004	1	629.53095964	26.95	0.0004

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	0.74069885	17.56	0.0001	0.04219084
GRADE	-0.04155546	-5.19	0.0004	0.00800421

EXPGR=1

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	204.71786370	204.71786370	26.01	0.0005	0.722261	733.9255
ERROR	10	78.72241277	7.87224128		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	11	283.44027647			2.80575146		0.38229376

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	204.71786370	26.01	0.0005	1	204.71786370	26.01	0.0005

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	0.65450711	13.05	0.0001	0.05014393
GRADE	-0.04448014	-5.10	0.0005	0.00872243

TABLE F6

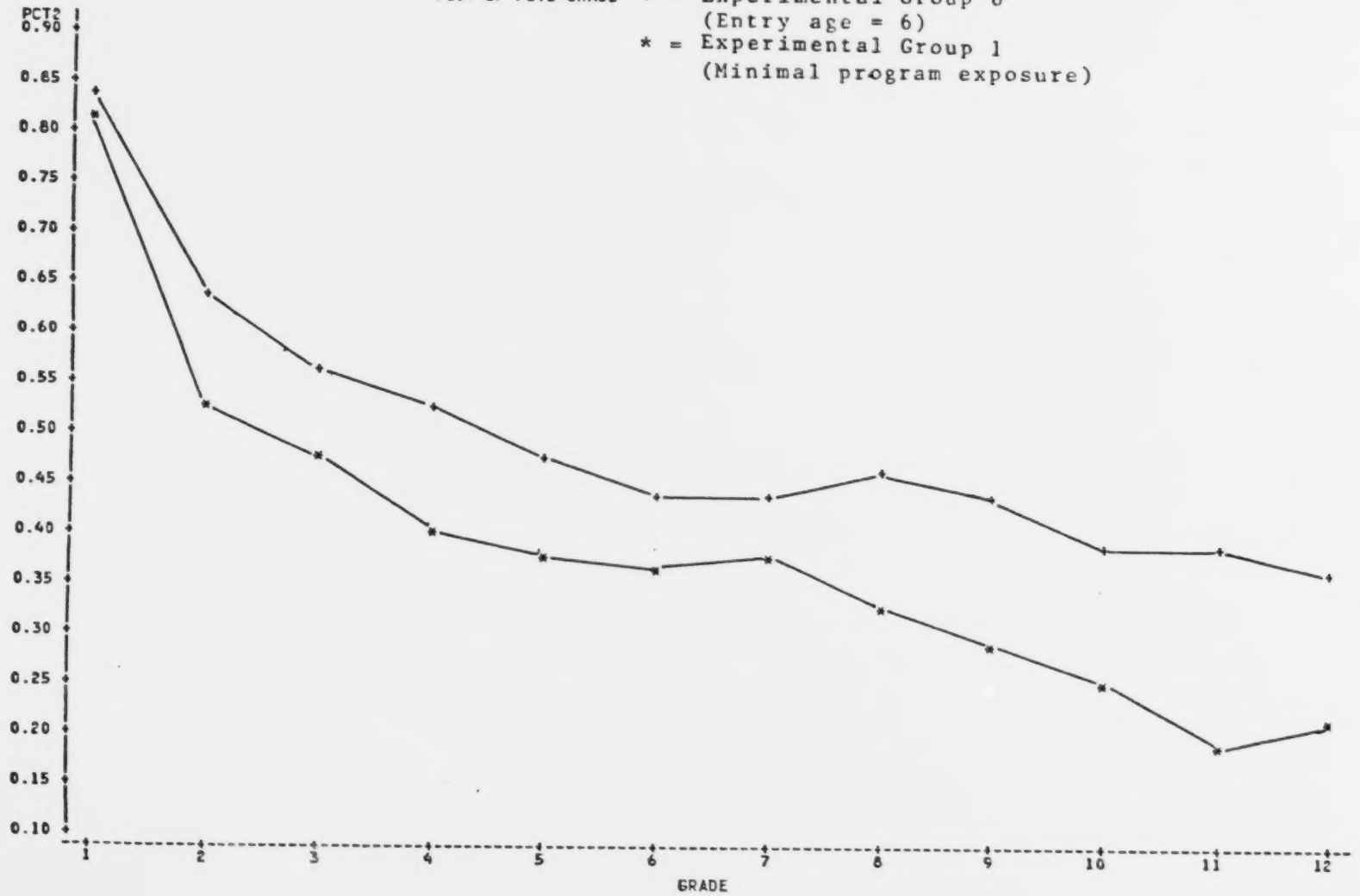
MEAN AGES OF STUDENTS USED IN EXPERIMENTAL GROUPS

VARIABLE	MEAN AGE
Exp group	
6	12.80
1	12.81
Grade level	
1	8.06
2	8.91
3	9.57
4	10.82
5	11.52
6	12.67
7	13.66
8	14.57
9	15.73
10	16.68
11	17.48
12	18.42

TABLE F7

$ASEXGL5

PLOT OF PCT2*GRADE + = Experimental Group 6
(Entry age = 6)
* = Experimental Group 1
(Minimal program exposure)



PLOT F1

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2 % AT OR ABOVE GRADE LEVEL
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	52375363.99830620	26187681.99915310	359.82	0.0001	0.785085	628.6109
ERROR	197	14337578.74748990	72779.58755071		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	199	66712942.74579600			269.77692183		42.91635962

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	25455874.75764560	349.77	0.0001	1	2144704.56603314	29.47	0.0001
GRADE	1	26919469.24066040	369.86	0.0001	1	26919469.24066040	369.86	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	76.25835979	67.65	0.0001	1.12781982
ENTR AGE	-0.69726843	-5.43	0.0001	0.12836575
GRADE	-3.70207028	-19.23	0.0001	0.19249349

MODEL G1

TABLE G1

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2 % AT OR ABOVE GRADE LEVEL
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	53678891.62964950	17892963.87654980	269.07	0.0001	0.804625	600.8815
ERROR	196	13034051.11614660	66500.86079667		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	199	66712942.74579600			257.87644483		42.91635962

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	25455874.75764560	382.79	0.0001	1	2860265.34166639	43.02	0.0001
GRADE	1	26919469.24066040	404.80	0.0001	1	13054229.43621640	196.30	0.0001
ENTR AGE*GRADE	1	1303527.63134329	19.60	0.0001	1	1303527.63134329	19.60	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	84.37521394	39.68	0.0001	2.12651704
ENTR AGE	-2.53046990	-6.56	0.0001	0.38605702
GRADE	-5.08831637	-14.01	0.0001	0.36317030
ENTR AGE*GRADE	0.20342567	4.43	0.0001	0.04594656

MODEL G2

TABLE G2

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2 % AT OR ABOVE GRADE LEVEL
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	60521817.23364570	12104363.44672910	379.29	0.0001	0.907198	416.2565
ERROR	194	6191125.51215035	31913.01810387		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	199	66712942.74579600			178.64215097		42.91635962

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
ENTR AGE	1	25455874.75764580	797.66	0.0001	1	47221.10113710	1.48	0.2253
GRADE	1	26919489.24066040	843.53	0.0001	1	17355403.34454040	543.83	0.0001
ENTR AGE*ENTR AGE	1	3062.64783403	0.10	0.7571	1	317663.32007564	9.95	0.0019
GRADE*GRADE	1	8124376.56693965	254.60	0.0001	1	5206115.62703537	163.16	0.0001
ENTR AGE*GRADE	1	19014.22056583	0.60	0.4411	1	19014.22056583	0.60	0.4411

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	85.97278796	56.56	0.0001	1.52013018
ENTR AGE	0.43655627	1.22	0.2253	0.36052983
GRADE	-10.33463803	-23.32	0.0001	0.44316133
ENTR AGE*ENTR AGE	-0.10726829	-3.16	0.0019	0.03399944
GRADE*GRADE	0.56920262	12.80	0.0001	0.04447965
ENTR AGE*GRADE	0.04663902	0.77	0.4411	0.05964455

MODEL G3

TABLE G3

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: PCT2 % AT OR ABOVE GRADE LEVEL
WEIGHT: TNUM

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	60459248.87373790	20153082.95791260	631.63	0.0001	0.906260	416.2147
ERROR	196	6253693.87205816	31906.60138805		ROOT MSE		PCT2 MEAN
CORRECTED TOTAL	199	66712942.74579600			178.62419038		42.91635962

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	50030578.43227300	1574.30	0.0001	1	18918029.92546250	592.73	0.0001
GRADE*GRADE	1	7155596.11366866	224.24	0.0001	1	8343369.24797389	261.46	0.0001
ENTR AGE*ENTR AGE	1	2873071.32776597	90.05	0.0001	1	2873071.32776597	90.05	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	86.69983383	80.34	0.0001	1.07942025
GRADE	-10.12465464	-24.35	0.0001	0.41567430
GRADE*GRADE	0.57724090	16.17	0.0001	0.03572750
ENTR AGE*ENTR AGE	-0.06288269	-9.49	0.0001	0.00662673

MODEL G4

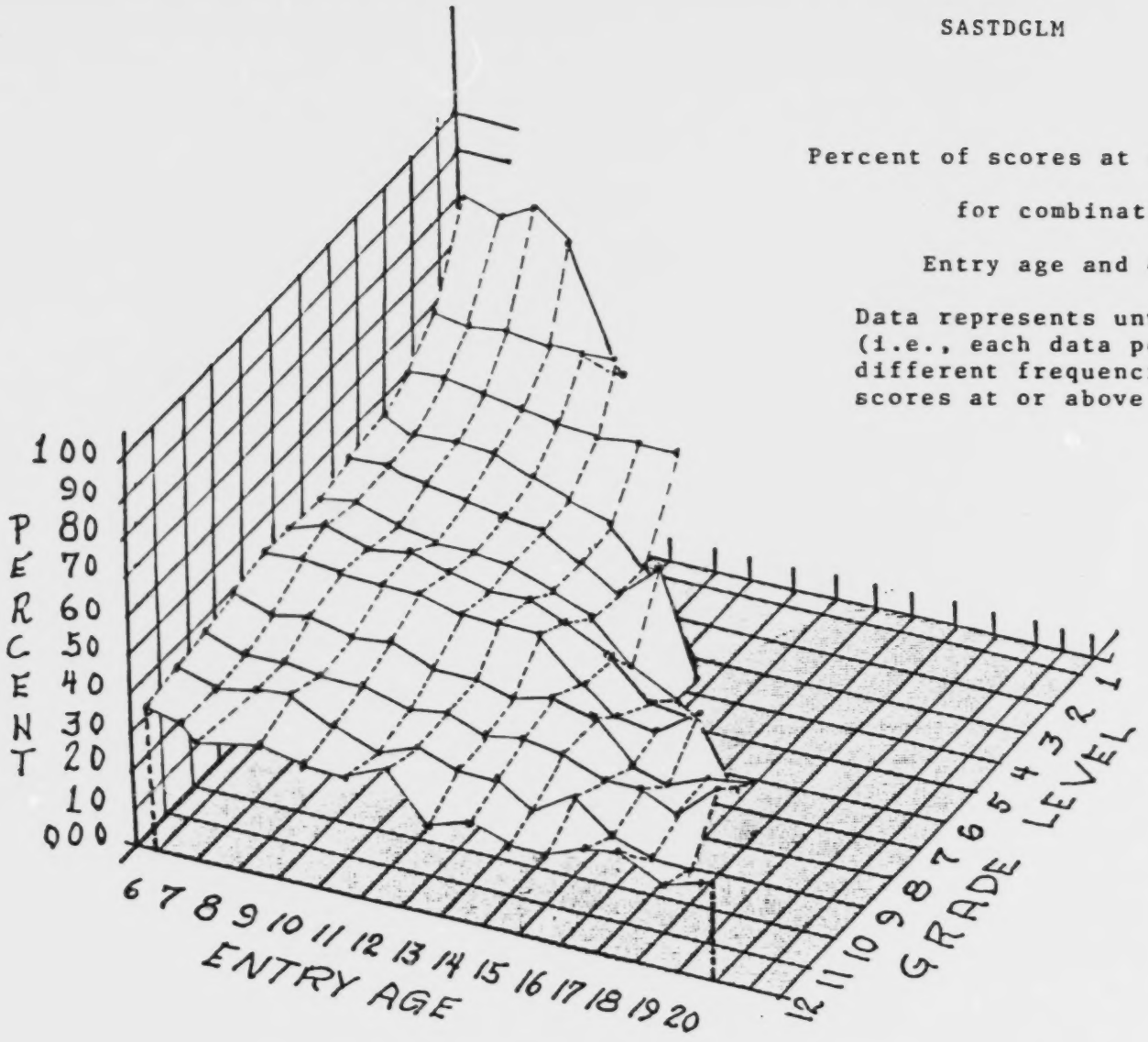
TABLE G4

Percent of scores at or above grade level

for combination of

Entry age and Grade level

Data represents unweighted percents (i.e., each data point represents different frequencies of percent of scores at or above grade level).



PLOT G1

ANALYSIS SUMMARY TABLE

Data Set : Skills
Cardinality: 406,559
Model : MEAN NUMBER OF MATH SKILLS = GRADE LEVEL

TYPE OF ANALYSIS->	ANOVA		REGRESSION				
			LINEAR		POLYNOMIAL		
TERM	-> P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
MATH_C	S	.0008	.0645	.6266	N	N	.6293
MATH_C OR READ_C	S	.0042	N	.2181	+.3810	_.0470	.7855
MATH_C AND READ_C	S	.0009	.0550	.3933	N	N	.4101

S = significant N = not significant +,- = direction of (significant) term

TABLE H1

* *

Data Set : SKILLS
Cardinality: 406,559
Population : MATH_C. (All records having at least one math skill).
Cardinality: 337,478
Source : SKGL4 18:35

Grade Level	number of records	number of skills	mean skills/ student	std. dev.
1	50,534	359,577	7.11	5.82
2	48,686	344,319	7.07	5.67
3	48,074	339,886	7.07	5.84
4	46,035	330,049	7.17	6.30
5	43,497	316,687	7.28	6.83
6	38,474	280,216	7.28	7.29
7	28,245	216,160	7.65	8.42
8	21,997	167,566	7.61	8.40
9	6,559	47,590	7.25	8.36
10	3,194	22,867	7.15	9.10
11	1,531	11,145	7.28	7.75
12	652	5,369	8.23	10.90
	337,478	2,441,431	7.23	6.72

TABLE H2

Data Set : SKILLS
Cardinality: 406,559
Population : MATH_C OR READ_C. (All records having either a math skill or a reading skill or both. In effect, this is the entire Skills data set .
Cardinality: 406,559

Grade Level	number of records	number of skills	mean skills/ student	std. dev.
1	59,478	359,577	6.04	5.93
2	57,143	344,319	6.02	5.81
3	55,887	339,886	6.08	5.95
4	53,754	330,049	6.13	6.81
5	50,789	316,687	6.23	6.81
6	45,341	280,216	6.18	7.21
7	35,344	216,160	6.11	8.13
8	28,122	167,566	5.95	8.06
9	10,210	47,590	4.66	7.55
10	5,626	22,867	4.06	7.72
11	3,221	11,145	3.46	6.46
12	1,635	5,369	3.28	7.97
	406,559	2,441,431	6.00	6.69

TABLE H3

* *

Data Set : SKILLS
Cardinality: 406,559
Population : MATH_C AND READ_C. (All records having both math skill(s) and reading skill(s).
Cardinality: 303,683

Grade Level	number of records	number of skills	mean skills/ student	std. dev.
1	46,666	341,233	7.31	5.92
2	44,658	324,925	7.27	5.78
3	43,770	318,024	7.26	5.95
4	41,917	308,150	7.35	6.36
5	39,510	295,493	7.47	7.00
6	34,778	260,082	7.47	7.49
7	24,615	193,297	7.85	8.36
8	18,920	148,951	7.87	8.62
9	4,872	34,941	7.17	7.19
10	2,350	16,214	6.90	6.98
11	1,134	7,933	6.99	6.71
12	483	3,518	7.28	8.34
	303,683	2,252,761	7.41	6.75

TABLE H4

GLM ANOVA OF NUMBER OF MATH SKILLS X GRADE LEVEL
(MATH SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: MATH # OF MATH SKILLS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	12529.08088453	1139.00735314	25.15	0.0001	0.000819	93.0192
ERROR	337466	15281745.04519480	45.28360650		ROOT MSE		MATH MEAN
CORRECTED TOTAL	337477	15294274.12607930			6.72932437		7.23434120

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	11	12529.08088452	25.15	0.0001	11	12529.08088452	25.15	0.0001

TABLE H5 SASSKGL4 18:35 6/5

GLM LINEAR REGRESSION OF MEAN NUMBER OF MATH SKILLS ON GRADE LEVEL
(MATH SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: M SKILL MEAN # OF MATH SKILLS PER GRADE
WEIGHT: NSTU

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	7850.88363859	7850.88363859	16.78	0.0022	0.626613	294.3032
ERROR	10	4678.19724595	467.81972460		ROOT MSE		M SKILL MEAN
CORRECTED TOTAL	11	12529.08088455			21.62914063		7.34927129

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.96393644	91.09	0.0001	0.07578413
GRADE	0.06454071	4.10	0.0022	0.01575483

TABLE H6 SASSKGL3 8:46 5/12

GLM POLYNOMIAL REGRESSION OF MEAN NUMBER OF MATH SKILLS ON GRADE LEVEL
(MATH SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

WEIGHT: MATH N

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	7884.55346044	3942.27673022	7.64	0.0115	0.629300	309.1044
ERROR	9	4644.52742410	516.05860268		ROOT MSE		MATH M MEAN
CORRECTED TOTAL	11	12529.08088455			22.71692327		7.34927129

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	7850.88363859	15.21	0.0036	1	343.82674503	0.67	0.4354
GRADE*GRADE	1	33.66982185	0.07	0.8041	1	33.66982185	0.07	0.8041

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.98993484	54.10	0.0001	0.12921008
GRADE	0.04960439	0.67	0.4354	0.06077148
GRADE*GRADE	0.00158094	0.26	0.8041	0.00616932

TABLE H7 SASTSG12 8:18 7/7

GLM ANOVA OF NUMBER OF MATH SKILLS X GRADE LEVEL
(MATH SKILLS OR READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: MATH		# OF MATH SKILLS					
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	78578.20320613	7143.47301874	159.42	0.0001	0.004295	111.4704
ERROR	406547	18216782.18596250	44.80855158		ROOT MSE		MATH MEAN
CORRECTED TOTAL	406558	18295360.38916860			6.69391899		6.00510873

TABLE H8 SASSKGL4 19:23 6/5

GLM LINEAR REGRESSION OF MEAN NUMBER OF MATH SKILLS ON GRADE LEVEL
(MATH SKILLS OR READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: M SKILL		MEAN # OF MATH SKILLS					
WEIGHT: NSTU							
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	17140.71817387	17140.71817387	2.79	0.1258	0.218136	1463.9187
ERROR	10	61437.48503228	6143.74850323		ROOT MSE		M SKILL MEAN
CORRECTED TOTAL	11	78578.20320615			78.38206749		5.35426359

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.36334863	25.74	0.0001	0.24720639
GRADE	-0.08284444	-1.67	0.1258	0.04959816

TABLE H9 SASSKGL1 14:49 5/12

GLM POLYNOMIAL REGRESSION OF MEAN NUMBER OF MATH SKILLS ON GRADE LEVEL
(MATH SKILLS OR READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: M MEAN							
WEIGHT: NSTU							
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	61730.64132145	30865.32066073	16.49	0.0010	0.785595	808.0673
ERROR	9	16847.56188470	1871.95132052		ROOT MSE		M MEAN MEAN
CORRECTED TOTAL	11	78578.20320615			43.26605275		5.35426359

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	17140.71817387	9.16	0.0143	1	27779.72916394	14.84	0.0033
GRADE*GRADE	1	44589.92314758	23.82	0.0009	1	44589.92314758	23.82	0.0009

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	5.52556807	25.20	0.0001	0.21928465
GRADE	0.36101601	3.85	0.0036	0.09370675
GRADE*GRADE	-0.04701907	-4.88	0.0009	0.00963404

TABLE H10 SASSKGL8 13:15 7/24

GLM ANOVA OF NUMBER OF MATH SKILLS X GRADE LEVEL
(MATH SKILLS AND READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: MATH C # OF MATH CURRENT SKILLS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	12591.63802317	1144.69436574	25.09	0.0001	0.000908	91.0535
ERROR	303671	13854345.78506600	45.62288063		ROOT MSE		MATH C MEAN
CORRECTED TOTAL	303682	13866937.42308928			6.75447116		7.41813338

TABLE H11 SASSKGL7 10:56 7/24

GLM LINEAR REGRESSION OF MEAN NUMBER OF MATH SKILLS ON GRADE LEVEL
(MATH SKILLS AND READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: M MEAN MEAN # OF MATH SKILLS
WEIGHT: M NSTU

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	4952.33215659	4952.33215659	6.48	0.0291	0.393303	375.8898
ERROR	10	7639.30586661	763.93058666		ROOT MSE		M MEAN MEAN
CORRECTED TOTAL	11	12591.63802320			27.63929425		7.35304658

TABLE H12 SASSKGL8 13:15 7/24

GLM POLYNOMIAL REGRESSION OF MEAN NUMBER OF MATH SKILLS ON GRADE LEVEL
(MATH SKILLS AND READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: M MEAN MEAN # OF MATH SKILLS
WEIGHT: M NSTU

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	5164.63231662	2582.31615831	3.13	0.0930	0.410164	390.6774
ERROR	9	7427.00570658	825.22285629		ROOT MSE		M MEAN MEAN
CORRECTED TOTAL	11	12591.63802320			28.72669240		7.35304658

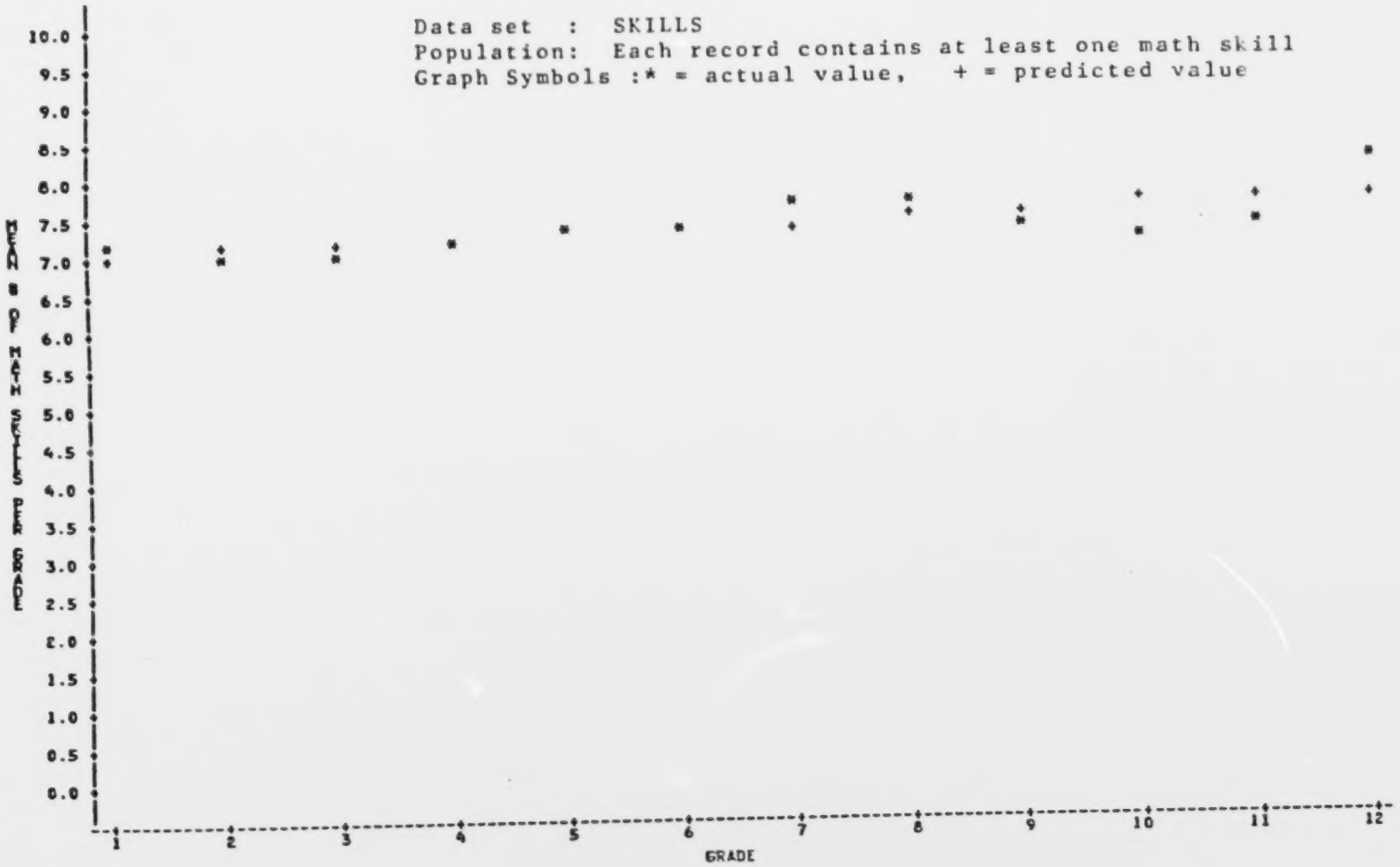
SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	4952.33215659	6.00	0.0368	1	1099.37251373	1.33	0.2760
GRADE*GRADE	1	212.30016003	0.26	0.6242	1	212.30016003	0.26	0.6242

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	7.13245040	41.27	0.0001	0.17256291
GRADE	0.09537205	1.15	0.2760	0.08259691
GRADE*GRADE	-0.00434454	-0.51	0.6242	0.00856552

TABLE H13 SASSKGL8 13:15 7/24

LINEAR REGRESSION OF MEAN NUMBER OF MATH SKILLS PER STUDENT ON GRADE LEVEL SASSKGL3

Data set : SKILLS
Population: Each record contains at least one math skill
Graph Symbols :* = actual value, + = predicted value

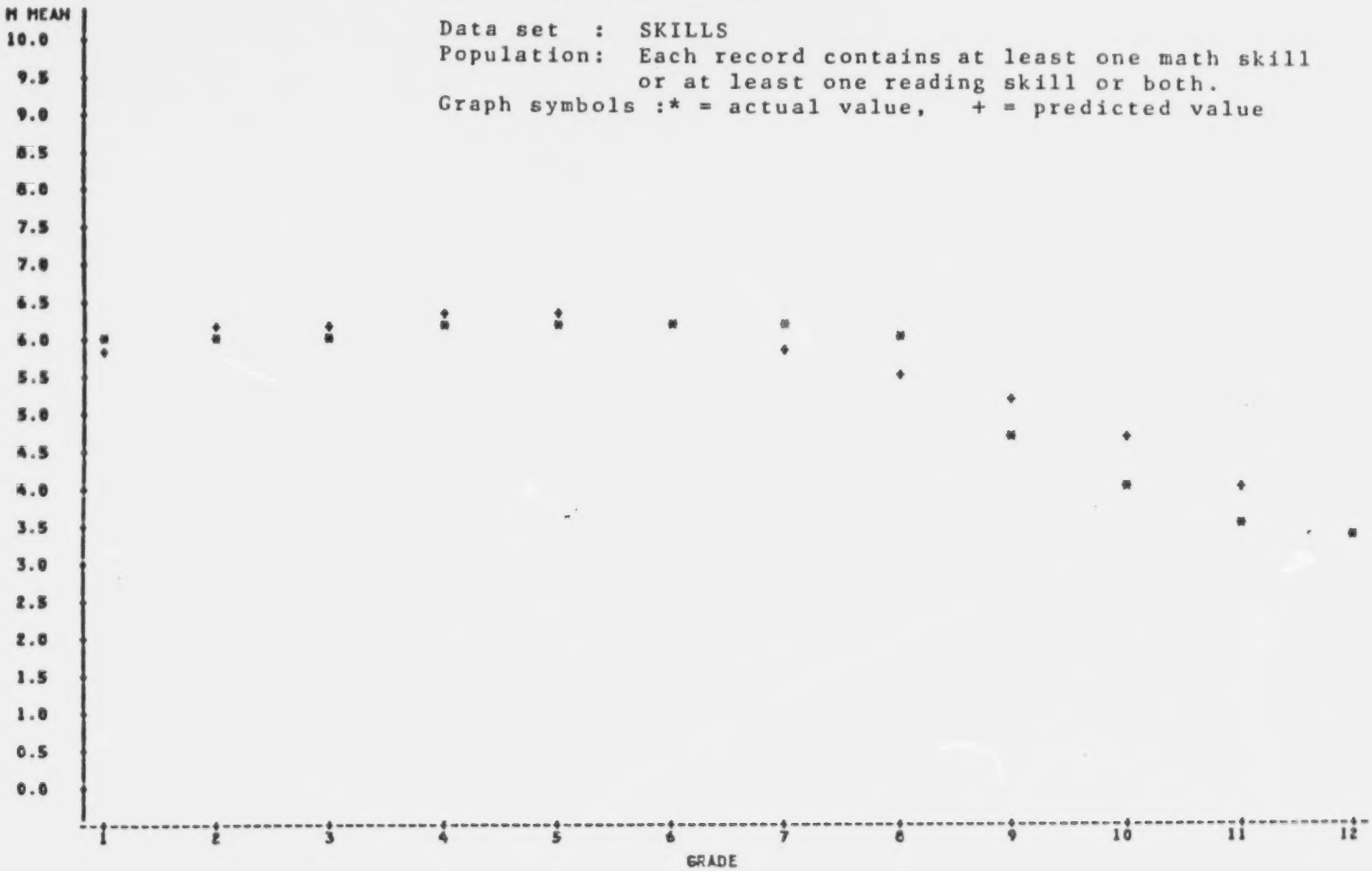


PLOT H1

POLYNOMIAL REGRESSION OF MEAN NUMBER OF MATH SKILLS PER STUDENT ON GRADE LEVEL

SASSKGL8

Data set : SKILLS
Population: Each record contains at least one math skill or at least one reading skill or both.
Graph symbols :* = actual value, + = predicted value



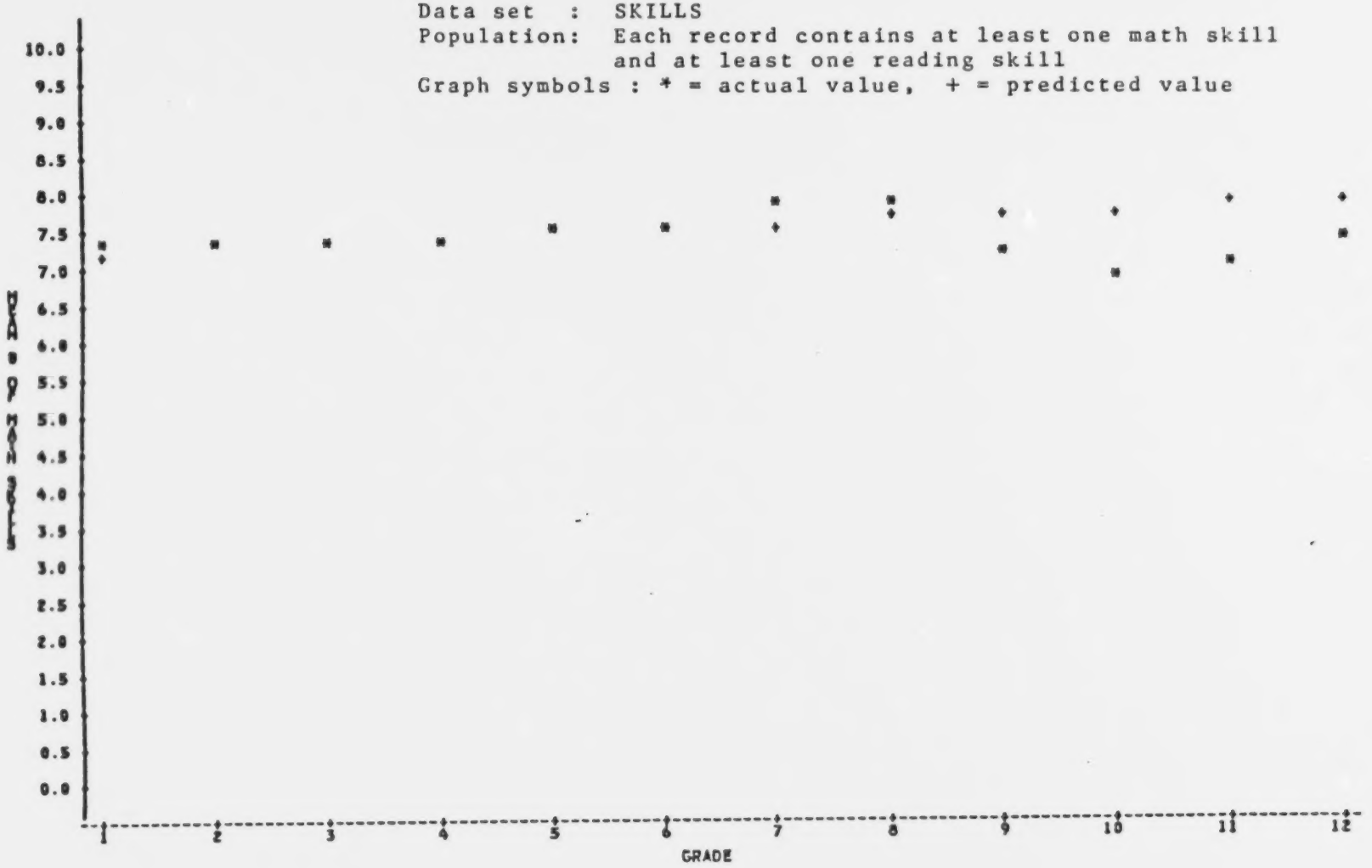
PLOT H2

LINEAR REGRESSION OF MEAN NUMBER
OF MATH SKILLS PER STUDENT
ON GRADE LEVEL

13:15 THURSDAY, JULY 24, 1986 3

SASSKGL8

Data set : SKILLS
Population: Each record contains at least one math skill
and at least one reading skill
Graph symbols : * = actual value, + = predicted value



PLOT H3

TABLE OF GRADE BY MATH TOPIC

GRADE / MATH TOPIC FREQUENCY PERCENT ROW PCT COL PCT	REDINESS	NUM MEAN	WHOL NUM	FRACTONS	DECIMALS	PERCENT	MEASURE	GEOMETRY	STATS/PR	SETS	TOTAL
1	50971 2.09 14.18 66.74	62546 2.56 17.39 55.87	206368 8.48 57.39 15.17	4969 0.20 1.38 1.20	191 0.01 0.05 0.16	52 0.00 0.01 0.07	31195 1.28 8.68 14.35	247 0.01 0.07 0.66	282 0.01 0.08 1.49	2756 0.11 0.77 24.56	359577 14.73
2	14933 0.61 4.34 19.55	26805 1.10 7.78 23.94	256305 10.50 74.44 18.84	8018 0.33 2.33 1.94	383 0.02 0.11 0.32	34 0.00 0.01 0.05	35831 1.47 10.41 16.46	563 0.02 0.16 1.50	434 0.02 0.13 2.30	1013 0.04 0.29 9.03	344319 14.10
3	6270 0.26 1.82 8.22	11329 0.46 3.29 10.11	272713 11.17 79.24 20.05	15275 0.63 4.44 3.69	1608 0.07 0.47 1.33	195 0.01 0.06 0.27	28850 1.18 8.38 13.27	1344 0.06 0.40 3.59	1456 0.06 0.43 7.70	843 0.03 0.25 7.51	339886 13.92
4	1713 0.07 0.52 2.24	4729 0.19 1.43 4.22	239911 9.83 72.69 17.64	48764 2.00 14.77 11.80	6114 0.25 1.85 5.05	399 0.02 0.12 0.54	22179 0.91 6.72 10.20	3209 0.13 0.97 8.58	2481 0.10 0.75 13.12	550 0.02 0.17 4.90	330049 13.52
5	1082 0.04 0.34 1.42	2897 0.12 0.91 2.59	172170 7.05 54.37 12.66	93389 3.83 29.49 22.59	17019 0.70 5.37 14.06	1619 0.07 0.51 2.21	21227 0.87 6.70 9.77	4333 0.18 1.37 11.58	2484 0.10 0.78 13.14	467 0.02 0.15 4.16	316687 12.97
6	859 0.04 0.31 1.12	1794 0.07 0.64 1.60	111206 4.55 39.69 8.17	99755 4.09 35.60 24.13	27496 1.13 9.81 22.72	8424 0.35 3.01 11.48	20858 0.85 7.44 9.60	5527 0.23 1.97 14.77	3401 0.14 1.21 17.99	896 0.04 0.32 7.99	280216 11.48
7	265 0.01 0.12 0.35	721 0.03 0.33 0.64	51961 2.13 24.04 3.82	71011 2.91 32.85 19.18	32456 1.33 15.01 26.82	25649 1.05 11.87 34.94	22899 0.94 10.59 10.53	7762 0.32 3.59 20.74	2458 0.10 1.14 13.00	978 0.04 0.45 8.72	216160 8.85
8	204 0.01 0.13 0.27	336 0.01 0.20 0.30	32614 1.34 19.46 2.40	49641 2.03 29.62 12.01	23434 0.96 13.98 19.36	23185 0.95 13.84 31.59	21739 0.89 12.97 10.00	11144 0.46 6.65 29.78	3622 0.15 2.16 19.16	1647 0.07 0.98 14.68	167566 6.86
9	31 0.00 0.07 0.04	404 0.02 0.85 0.36	10491 0.43 22.04 0.77	11941 0.49 25.09 2.89	7073 0.29 14.86 5.84	7400 0.30 15.55 10.08	6641 0.27 13.95 3.06	1295 0.05 2.72 3.46	1420 0.06 2.98 7.51	894 0.04 1.88 7.97	47590 1.95
10	20 0.00 0.09 0.03	206 0.01 0.90 0.18	4015 0.16 17.56 0.30	5870 0.24 25.67 1.42	3118 0.13 13.64 2.58	3931 0.16 17.19 5.36	3455 0.14 15.11 1.59	1077 0.04 4.71 2.88	531 0.02 2.81	644 0.03 2.82 5.74	22867 0.94
11	2 0.00 0.02 0.00	85 0.00 0.76 0.08	1722 0.07 15.45 0.13	3113 0.13 27.93 0.75	1436 0.06 12.88 1.19	1783 0.07 16.00 2.43	1771 0.07 15.89 0.81	674 0.03 6.05 1.80	192 0.01 1.72 1.02	367 0.02 3.29 3.27	11145 0.46
12	11 0.00 0.20 0.01	106 0.00 1.97 0.09	886 0.04 16.50 0.07	1663 0.07 30.97 0.40	687 0.03 12.80 0.57	734 0.03 13.67 1.00	725 0.03 13.50 0.33	247 0.01 4.60 0.66	145 0.01 2.70 0.77	165 0.01 3.07 1.47	5369 0.22
TOTAL	76369 3.13	111953 4.59	1360362 55.72	413409 16.93	121015 4.96	73405 3.01	217370 8.90	37422 1.53	18906 0.77	11220 0.46	2441431 100.00

TABLE I1

SUMMARY PARAMETERS OF THE SKILLS DATA SET AND THE SKILLS/TEST DATA SET

From the MSRTS data base the below data sets were created. The SKILLS data set contains 406,559 grade 1-12 skill records. Each record contains at least one math skill or reading skill or both. The TEST Data Set contains 351,011 records each of which contains data on at least one math test or one reading test or both.

MATH SKILL CONTENT IN THE SKILLS Data Set		MATH TEST CONTENT IN THE TEST FILE	
number of records /w math skills	number of math skills	number of records in IDSHTD	number of math tests
337,478	2,441,431	351,011	139,680

The above two data sets (including the reading skills and reading tests) were integrated into the combined SKILL/TEST data set whose Math Skill and Math test content is shown below.

MATH SKILL AND MATH TEST CONTENT OF THE SKILLS/TEST DATA SET (IDSHTDSK)

record contents	number of records	number of skills	number of tests
tests with no skills	65,134	0	67,637
skills with no matching tests	4,626	31,371	0
skills with matching tests	66,734	602,216	72,043
total	136,494	633,587	139,680
total math tests in file	131,868		139,680
total math skills in file	71,360	633,587	

Compiled from TSM11 7:25 6/6

TABLE J1

ANALYSIS SUMMARY TABLE

Data Set : Skills/Test (IDSHTDSK)
Cardinality: 144,148
Model : MEAN NUMBER OF MATH SKILLS = GRADE LEVEL

TYPE OF ANALYSIS-> ANOVA			REGRESSION				
			LINEAR		POLYNOMIAL		
TERM -> ->	P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
1 MATH_C	S	.0018	N	.0145	N	N	.0163
2 MATH_C OR READ_C	S	.0242	N	.2524	N	-.0734	.6294
3 MATH_C AND READ_C	S	.0018	N	.1077	N	N	.1313
4 MATH_T	S	.0466	-.4653	.6719	N	-.0852	.9045
5 MATH_T OR READ_T	S	.0478	-.4577	.7047	N	-.0804	.9055
6 MATH_T AND READ_T	S	.0460	-.4649	.6867	+.5108	-.0861	.9037
7 MATH_C AND MATH_T	S	.0020	N	.0282	N	N	.0288
8 MATH_C OR READ_C AND MATH_T	S	.0107	N	.2259	+.6329	-.0820	.6245

S = significant N = not significant +,- = direction of (significant) term

TABLE J2

****

ANALYSIS SUMMARY TABLE

Data Set : Skills/Test (IDSHTDSK)
Cardinality: 144,148
Model : MEAN NUMBER OF MATH TESTS = GRADE LEVEL

TYPE OF ANALYSIS-> ANOVA			REGRESSION				
			LINEAR		POLYNOMIAL		
TERM -> ->	P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
1 MATH_C	S	.0015	N	.0175	+.0177	-.0019	.6038
2 MATH_C OR READ_C	S	.0012	N	.0551	N	-.0012	.4095
3 MATH_C AND READ_C	S	.0016	N	.0269	+.0183	-.0021	.5955
4 MATH_T	S	.0024	-.0041	.9222	N	N	.9340
5 MATH_T OR READ_T	S	.0078	-.0080	.4912	+.0207	-.0025	.9280
6 MATH_T AND READ_T	S	.0024	-.0041	.9098	N	N	.9194
7 MATH_C AND MATH_T	S	.0006	N	.0666	-.0101	+.0010	.7062
8 MATH_C OR READ_C AND MATH_T	S	.0006	N	.2436	-.0078	+.0006	.6030

S = significant N = not significant +,- = direction of (significant) term

TABLE J3

ANALYSIS SUMMARY TABLE

Data Set : Skills/Test (IDSHTDSK)
Cardinality: 144,148
Model : MEAN NUMBER OF MATH TEST = MATH SKILLS PER STUDENT

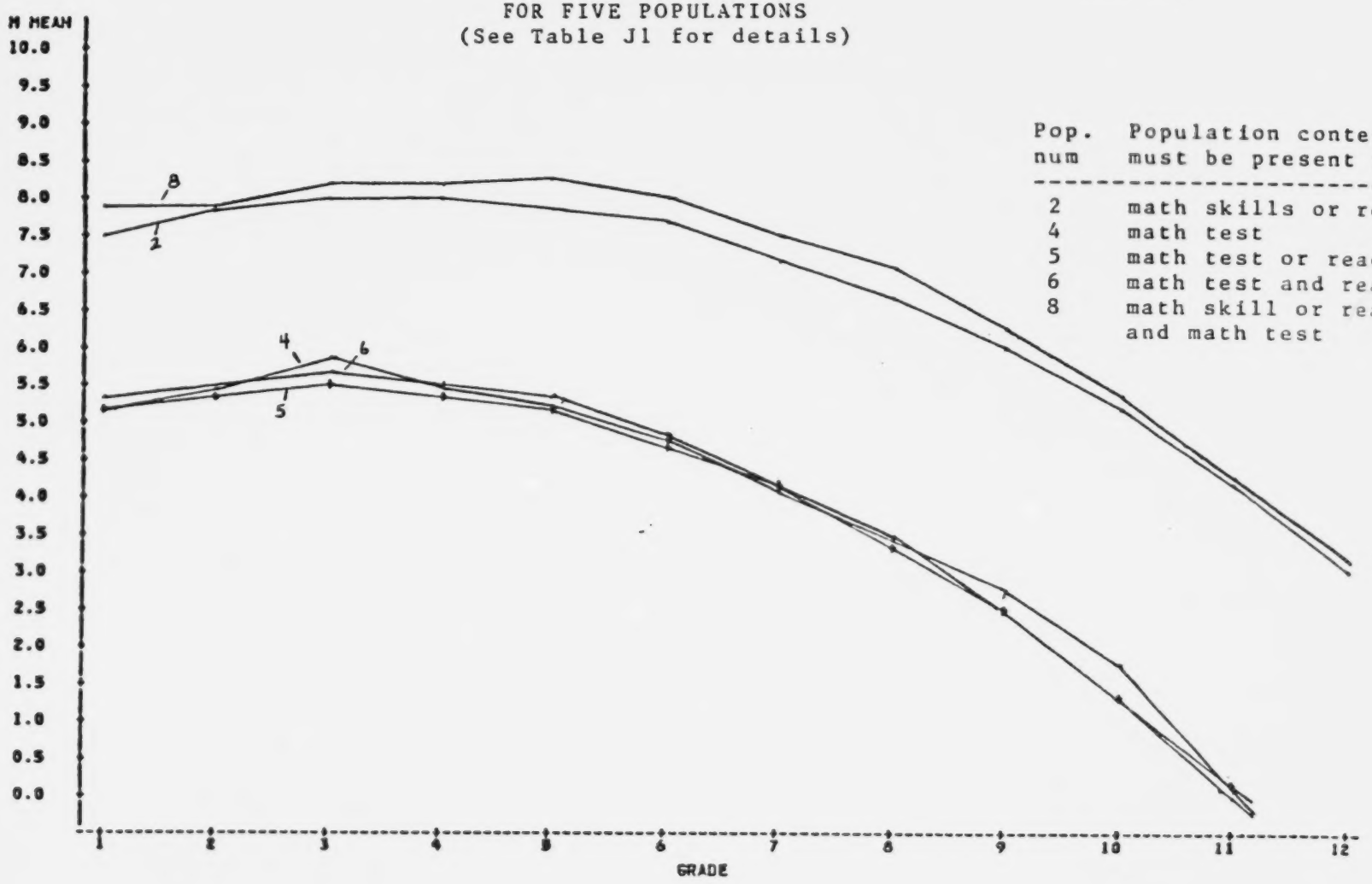
TYPE OF ANALYSIS->	ANOVA		REGRESSION				
			LINEAR		POLYNOMIAL		
TERM -> ->	P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
1 MATH_C	S	.0132	.0050	.6662	N	N	.6966
2 MATH_C OR READ_C	S	.0407	.0081	.5333	+.0144	N	.5904
3 MATH_C AND READ_C	S	.0154	.0051	.6378	N	N	.6682
4 MATH_T	S	.0126	.0007	.6292	N	N	.6330
5 MATH_T OR READ_T	S	.0180	.0068	.7930	+.0087	N	.8063
6 MATH_T AND READ_T	S	.0068	.0032	.6133	N	N	.6213
7 MATH_C AND MATH_T	S	.0082	.0021	.3796	N	+.?	.6874
8 MATH_C OR READ_C AND MATH_T	S	.0091	.0024	.4597	N	N	.5930

S = significant N = not significant +,- = direction of (significant) term

TABLE J4

POLYNOMIAL REGRESSION OF THE MEAN
NUMBER OF MATH SKILLS PER STUDENT
ON GRADE LEVEL
FOR FIVE POPULATIONS
(See Table J1 for details)

SASTSG10



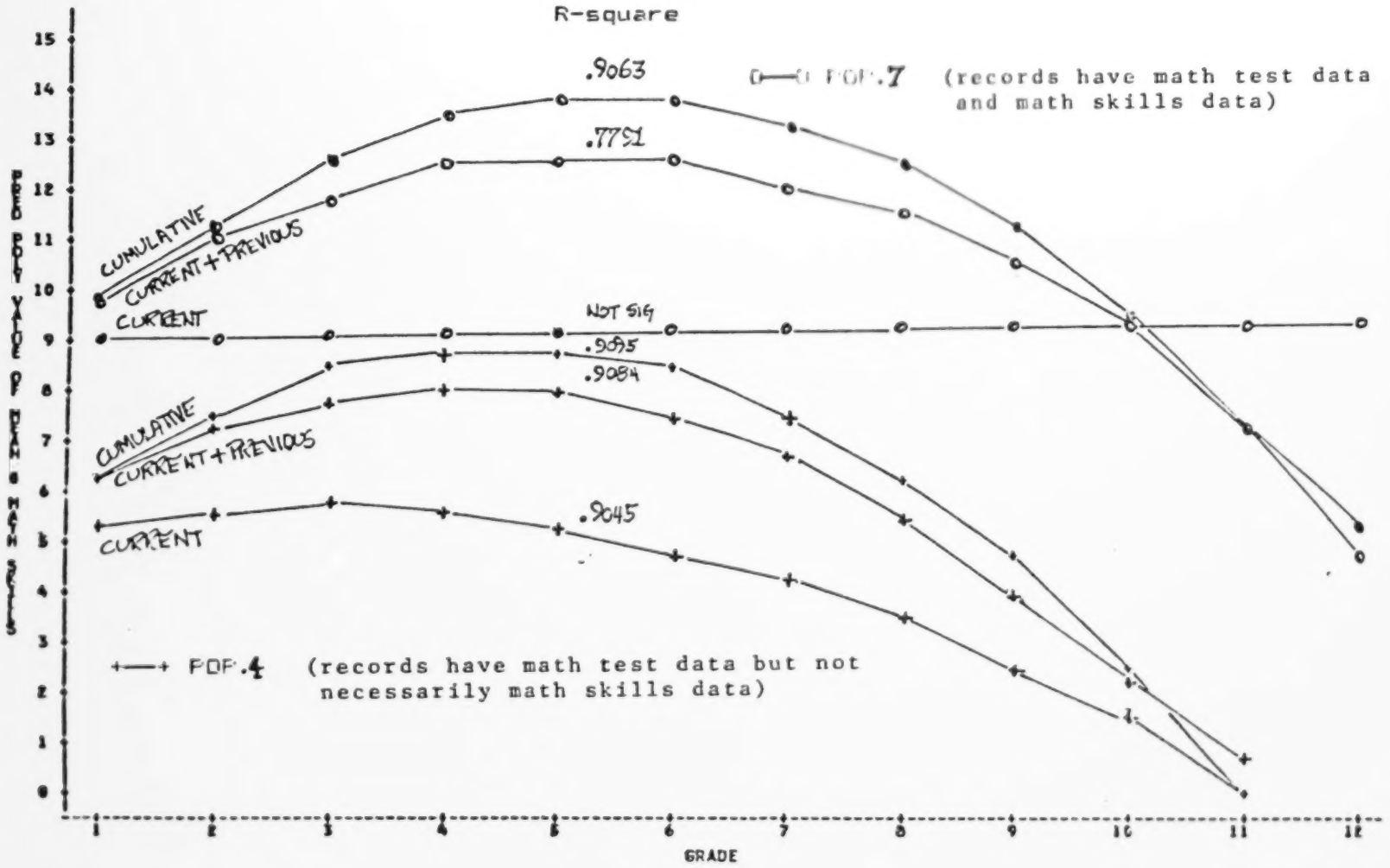
Pop. num	Population content that must be present in record
2	math skills or read skills
4	math test
5	math test or read test
6	math test and read test
8	math skill or read skill and math test

PLOT J1

BEST FIT POLYNOMIAL OF MEAN NUMBER OF CURRENT, CURRENT + PREVIOUS AND CUMULATIVE MATH SKILLS PER STUDENT

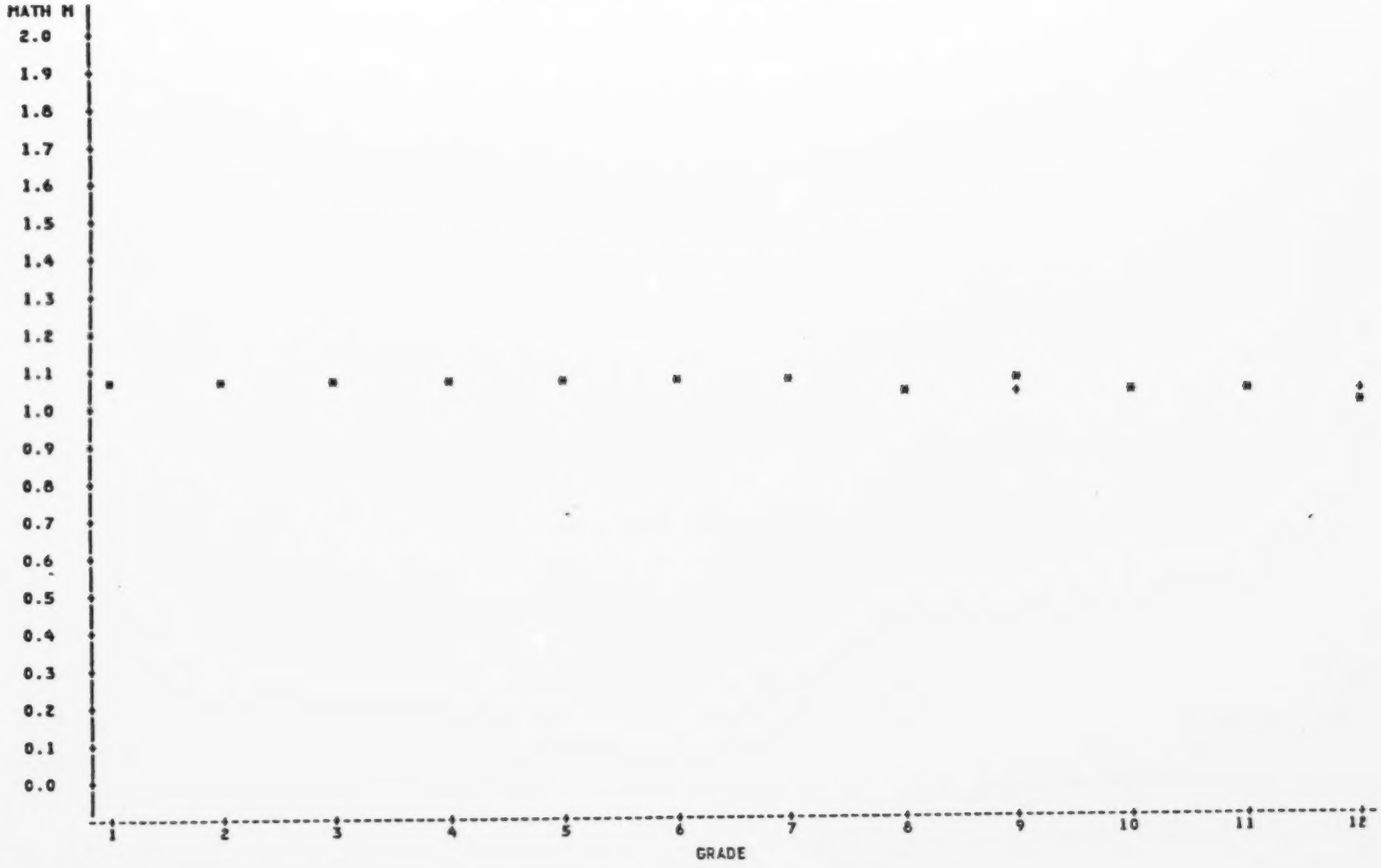
14:25 FRIDAY, JUNE 6, 1986 10

SASTSMN10.1



PLOT J2

PLOT OF MATH M*GRADE SYMBOL USED IS *
PLOT OF MATH P*GRADE SYMBOL USED IS +

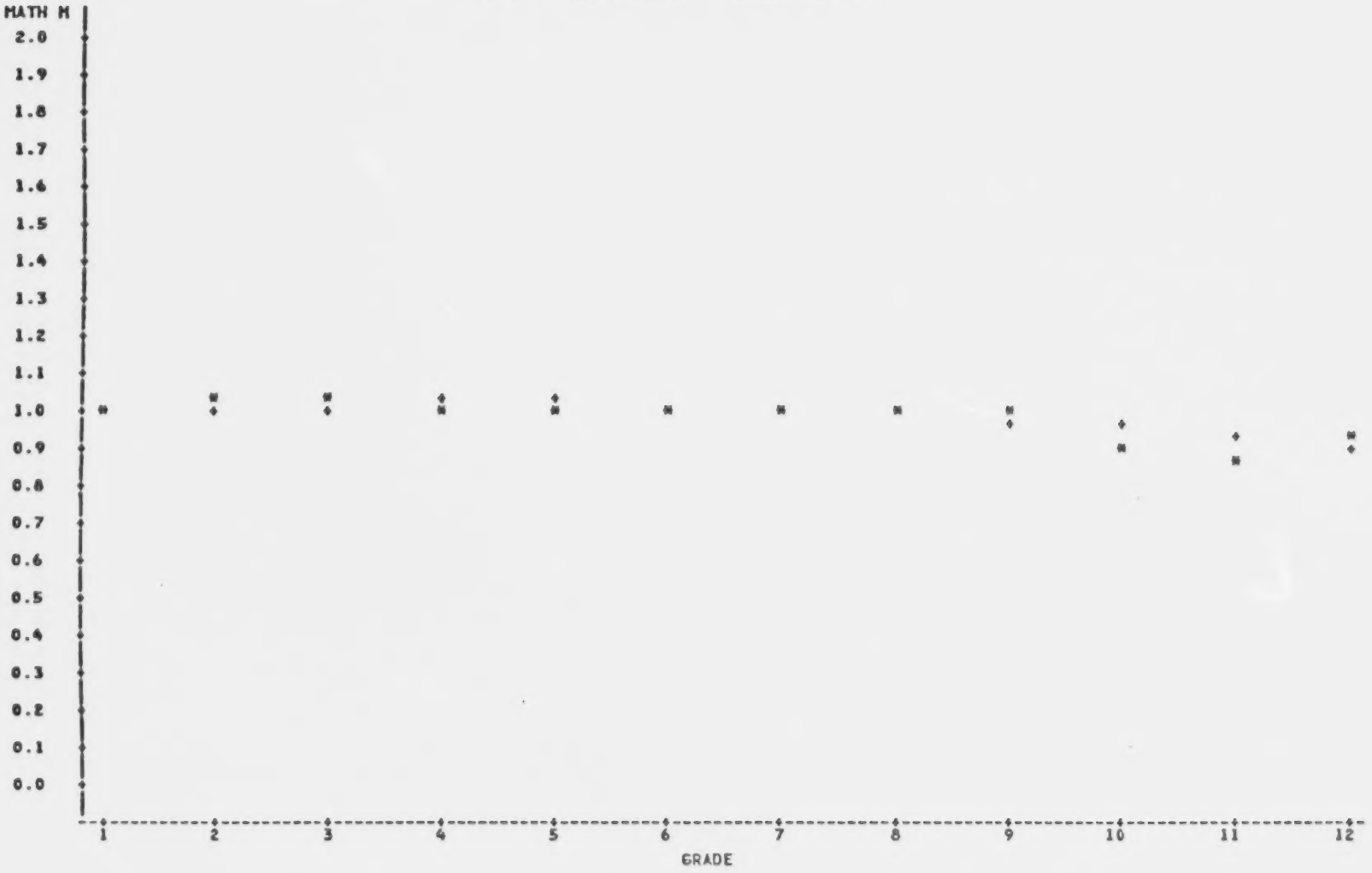


PLOT J3

242

243

PLOT OF MATH M*GRADE SYMBOL USED IS *
PLOT OF MATH P*GRADE SYMBOL USED IS +



PLOT J4

SUMMARY TABLES OF THREE POPULATIONS OF SCORES ON MATH TESTS AS A FUNCTION OF THE NUMBER OF MATH SKILLS IN STUDENTS RECORDS IN THE SKILLS/TEST DATA SET

C U R R E N T M A T H S K I L L S

	Type of Regression Model					
	Linear		Polynomial			
	slope	R-sq.	linear term	quad. term	R-sq	
GRADE						
01	0.19	.37	NS	NS	.37	
02	0.22	.59	.38	NS	.65	
03	NS	.19	NS	NS	.22	
04	0.15	.72	.20	NS	.73	POPULATION 1:
05	0.19	.46	NS	NS	.51	Records containing at least one
06	0.16	.29	NS	NS	.37	math test but not necessarily
07	NS	.16	NS	NS	.22	math skill data
08	0.37	.30	NS	NS	.31	
09	-1.29	.59	-2.21	NS	.68	
10	-1.83	.75	-3.54	.10	.90	TABLE K1
11	-1.46	.68	NS	NS	.68	
12	-1.86	.68	-3.63	.07	.82	
GRADE						
01	0.22	.33	NS	NS	.34	
02	0.22	.46	+0.53	NS	.60	
03	NS	.12	NS	NS	.17	
04	0.16	.65	0.29	NS	.71	POPULATION 2:
05	0.28	.62	NS	NS	.62	Records containing at least one
06	0.28	.57	NS	NS	.58	math skill and at least one
07	0.38	.54	0.77	NS	.61	math test.
08	0.68	.63	1.69	-0.03	.78	
09	NS	.14	NS	NS	.14	
10	-1.08	.36	-3.80	0.11	.66	TABLE K2
11	-1.67	.49	NS	NS	.52	
12	NS	.32	-4.24	N	.54	
GRADE						
01	0.19	.29	NS	NS	.30	
02	0.18	.39	NS	NS	.44	
03	NS	.04	NS	NS	.04	
04	0.13	.46	NS	NS	.46	POPULATION 3:
05	0.25	.58	NS	NS	.58	Records containing at least one
06	0.30	.63	0.43	NS	.65	math skill or one reading skill
07	0.38	.59	0.68	NS	.65	or both and one math test.
08	0.55	.50	NS	NS	.52	
09	NS	.18	NS	NS	.18	
10	-1.32	.60	-2.77	0.07	.77	TABLE K3
11	NS	.16	NS	NS	.24	
12	-1.39	.56	-2.66	NS	.66	

SUMMARY TABLES OF ONE POPULATION OF SCORES ON MATH TESTS AS A FUNCTION OF THE NUMBER OF MATH SKILLS IN STUDENTS RECORDS IN THE SKILLS/TEST DATA SET

CURRENT + PREVIOUS MATH SKILLS

	Type of Regression Model				
	Linear		Polynomial		
	slope	R-sq.	linear term	quad. term	R-sq
GRADE					
01	0.19	.36	NS	NS	.36
02	0.13	.55	NS	NS	.55
03	NS	.09	NS	NS	.10
04	0.09	.39	NS	.01	.60
05	0.15	.38	NS	.01	.69
06	0.20	.37	NS	NS	.56
07	NS	.14	-0.45	.02	.54
08	NS	.23	NS	.02	.58
09	-0.62	.50	-1.25	NS	.63
10	-1.41	.78	-2.25	.04	.85
11	-1.08	.73	-1.26	NS	.76
12	-1.64	.75	-2.50	NS	.80

POPULATION 1:
Records containing at least one math test but not necessarily math skill data.

TABLE K4

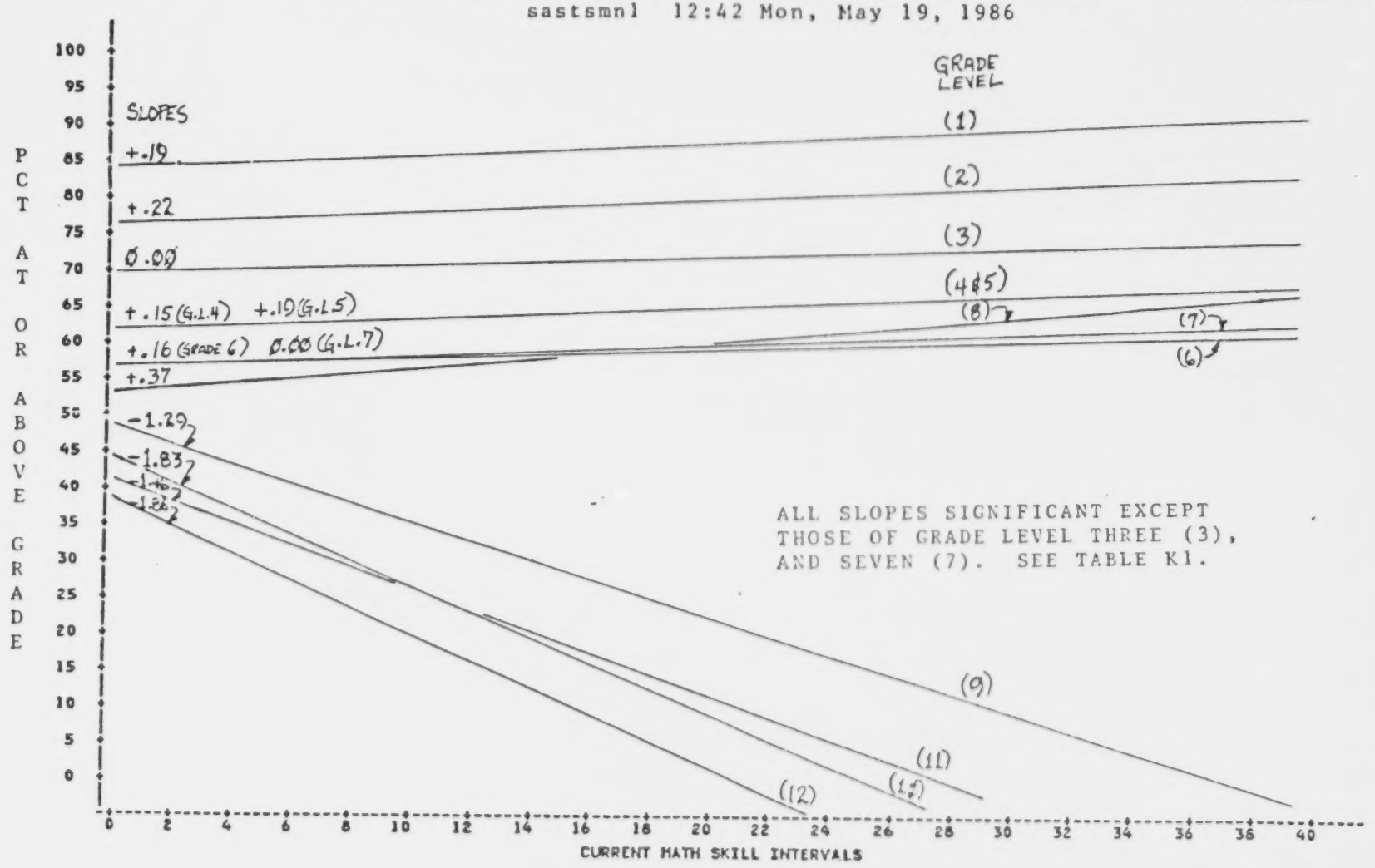
CUMULATIVE MATH SKILLS

	Type of Regression Model				
	Linear		Polynomial		
	slope	R-sq.	linear term	quad. term	R-sq
GRADE					
01	0.19	.36	NS	NS	.36
02	0.13	.54	NS	NS	.54
03	NS	.01	NS	NS	.09
04	NS	.09	NS	NS	.13
05	NS	.19	-0.24	0.01	.58
06	0.14	.38	NS	0.01	.70
07	NS	.10	NS	0.01	.44
08	NS	.19	-0.61	0.02	.62
09	-0.52	.51	-1.21	0.02	.69
10	-1.14	.84	-1.58	NS	.87
11	-0.76	.47	NS	NS	.48
12	-1.40	.70	-2.06	NS	.74

POPULATION 1:
Records having at least one math test but not necessarily math skill data.

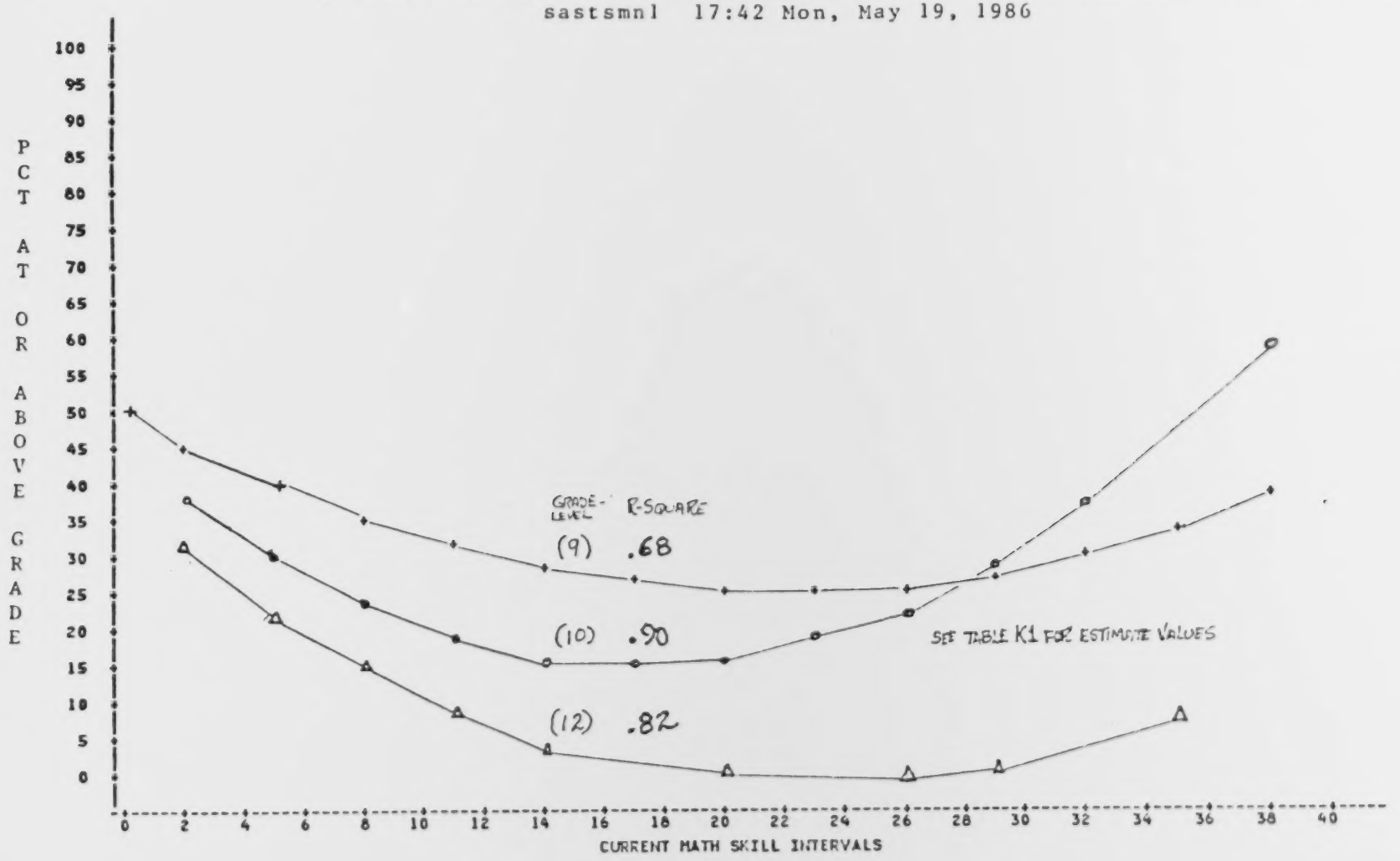
TABLE K5

LINEAR REGRESSION CURVES OF BEST FIT OF
PERCENT OF MATH SCORES AT OR ABOVE GRADE LEVEL
ON THE
NUMBER OF CURRENT MATH SKILLS IN STUDENT RECORDS FOR POPULATION 1:
(RECORDS HAVING AT LEAST ONE MATH TEST BUT NOT NECESSARILY MATH SKILLS DATA)
sastsmnl 12:42 Mon, May 19, 1986



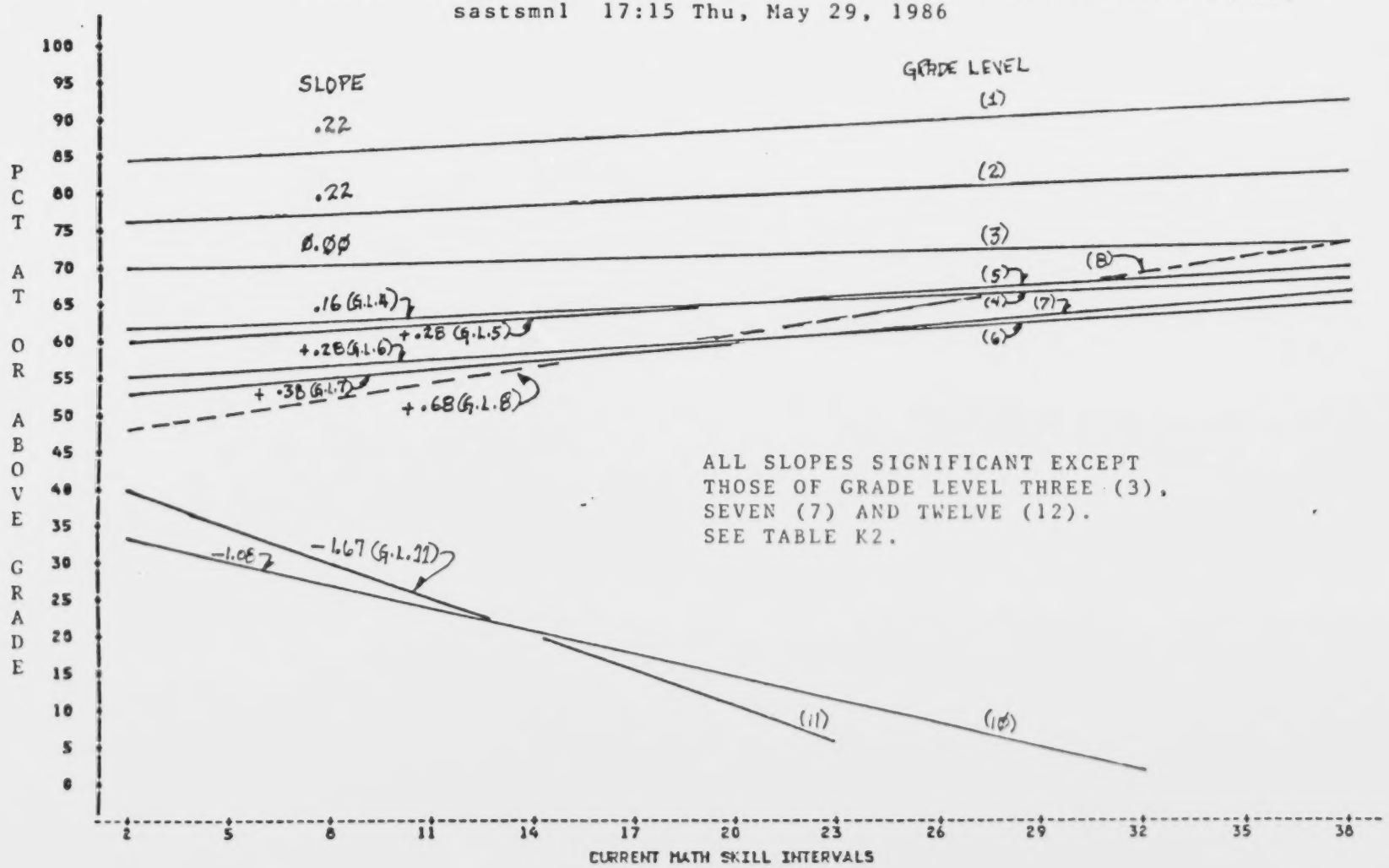
PLOT K1

POLYNOMIAL REGRESSION CURVES OF BEST FIT OF
PERCENT OF MATH SCORES AT OR ABOVE GRADE LEVEL
OF THE
NUBMER OF CURRENT MATH SKILLS IN STUDENT RECORDS FOR POPULATION 1:
(RECORDS HAVING AT LEAST ONE MATH TEST BUT NOT NECESSARILY MATH SKILLS DATA)
sastsmn1 17:42 Mon, May 19, 1986



PLOT K2

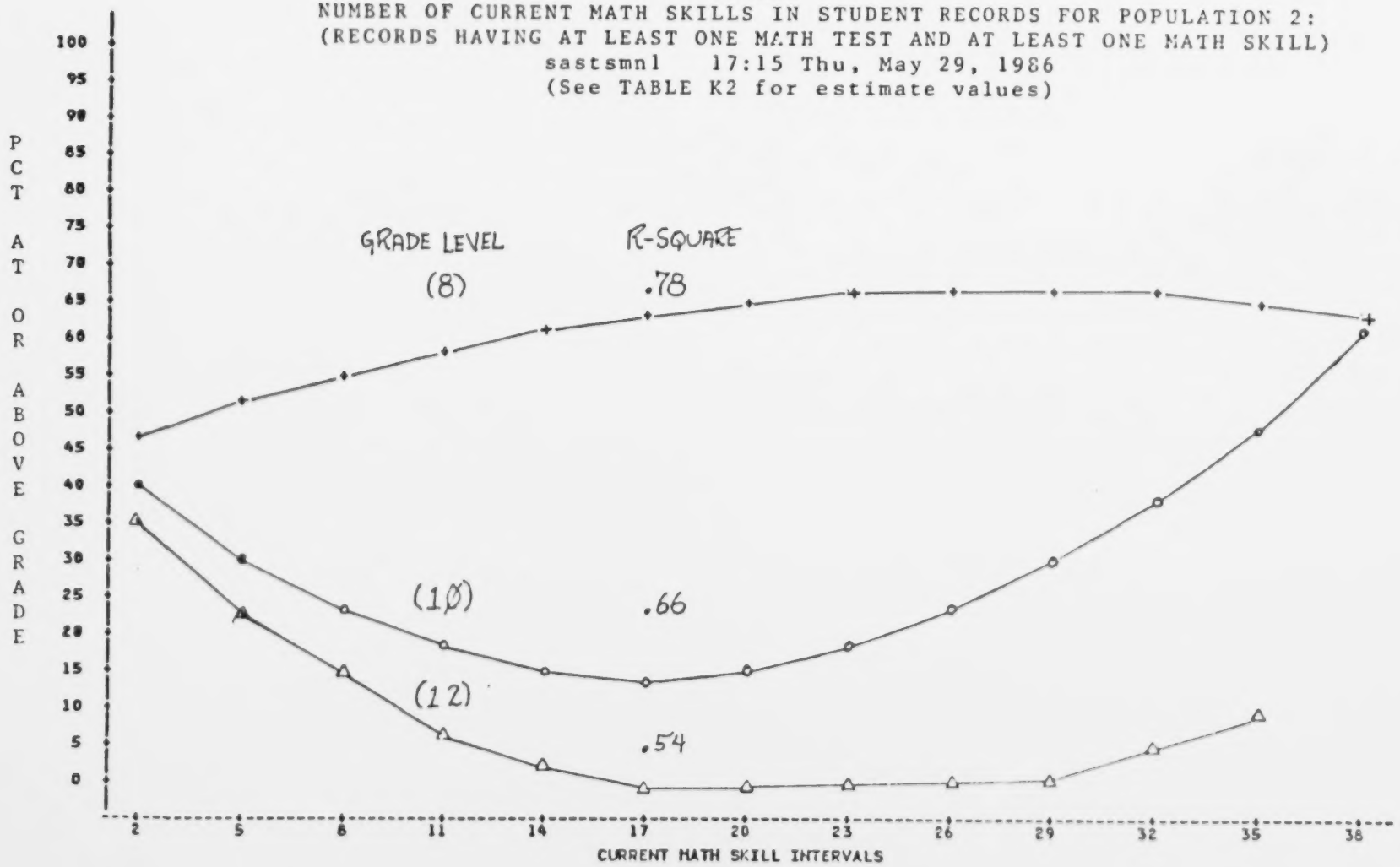
LINEAR REGRESSION CURVES OF BEST FIT OF
PERCENT OF MATH SCORES AT OR ABOVE GRADE LEVEL
ON THE
NUMBER OF CURRENT MATH SKILLS IN STUDENT RECORDS FOR POPULATION 2:
(RECORDS HAVING AT LEAST ONE MATH TEST AND AT LEAST ONE MATH SKILL)
sastsmnl 17:15 Thu, May 29, 1986



PLOT K3

59

POLYNOMIAL REGRESSION CURVES OF BEST FIT OR
PERCENT OF MATH SCORES AT OR ABOVE GRADE LEVEL
OF THE
NUMBER OF CURRENT MATH SKILLS IN STUDENT RECORDS FOR POPULATION 2:
(RECORDS HAVING AT LEAST ONE MATH TEST AND AT LEAST ONE MATH SKILL)
sastsmn1 17:15 Thu, May 29, 1986
(See TABLE K2 for estimate values)

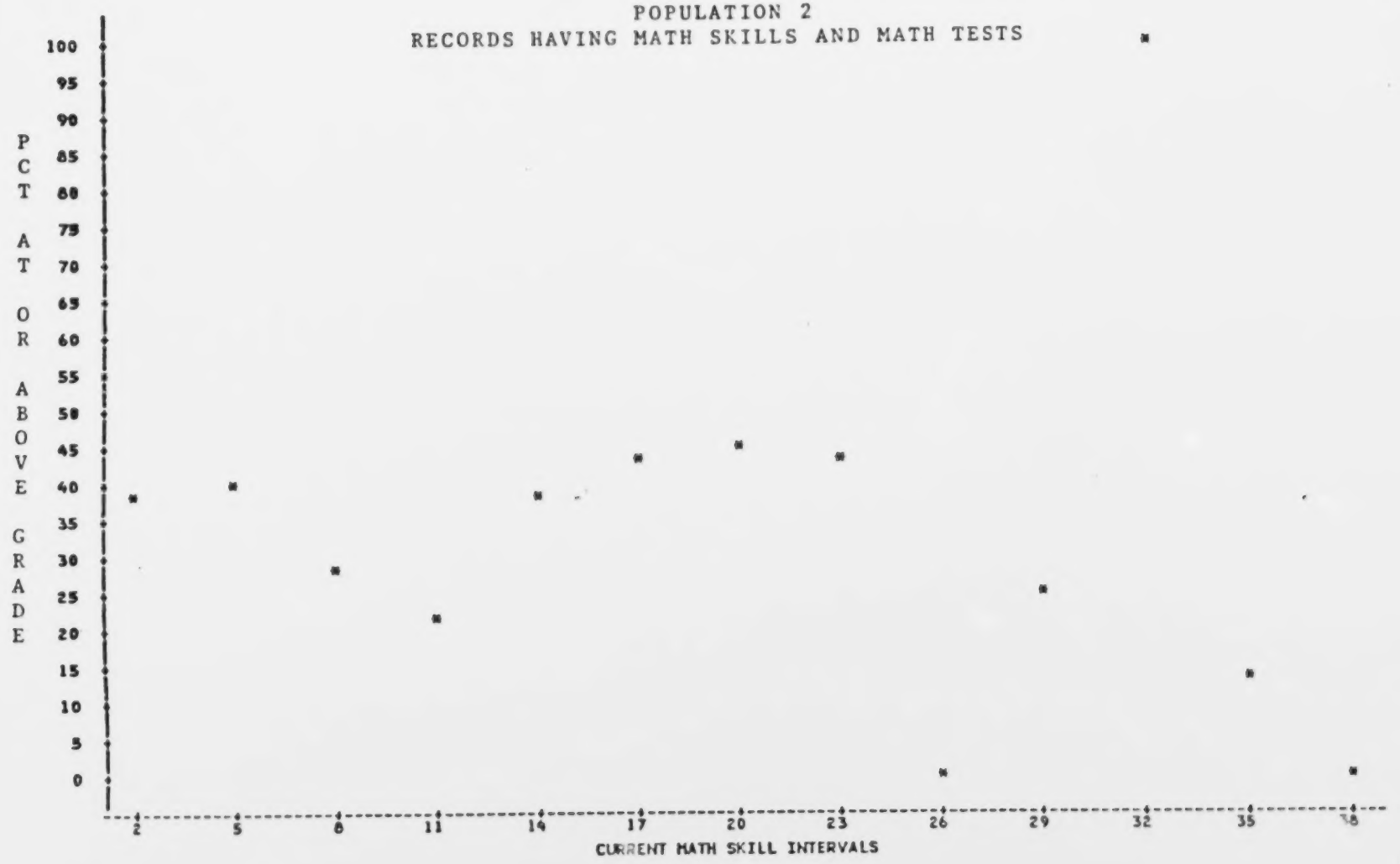


60

PLOT K4

GRADE=09

POPULATION 2
RECORDS HAVING MATH SKILLS AND MATH TESTS



PLOT K5

ANALYSIS SUMMARY TABLE

Data Set : Skills
Cardinality: 406,559
Model : MEAN NUMBER OF READING SKILLS = GRADE LEVEL

TYPE OF ANALYSIS->	ANOVA		REGRESSION				
			LINEAR		POLYNOMIAL		
TERM	-> P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
READ_C	S	.0078	-.1562	.9209	-.2972	+.0143	.9841
MATH_C OR READ_C	S	.0077	-.1656	.8463	-.2777	+.0100	.9874
MATH_C AND READ_C	S	.0099	-.1783	.9426	-.4048	+.0257	.9669

S = significant N = not significant +,- = direction of (significant) term

TABLE L1

* * * *

Data Set : SKILLS
Cardinality: 406,559
Population : READ_C OR MATH_C. (All records having at least one reading skill or at least one math skill).
Cardinality: 406,559
Source : SKGL4 18:35

Grade Level	number of records	number of skills	mean skills/ student	std. dev.
1	59,478	338,257	5.68	5.26
2	57,143	307,040	5.37	4.88
3	55,887	285,979	5.11	4.70
4	53,754	265,885	4.94	4.29
5	50,798	247,601	4.87	4.46
6	45,341	213,271	4.70	4.30
7	35,344	154,867	4.38	4.00
8	28,122	122,922	4.37	3.88
9	10,210	39,997	3.91	3.63
10	5,626	23,326	4.14	3.96
11	3,221	13,882	4.30	3.98
12	1,635	7,106	4.36	3.60
	406,559	2,020,133	4.96	4.54

TABLE L3

* * * \

GLM ANOVA OF NUMBER OF READING SKILLS X GRADE LEVEL
(READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: READ

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	59683.45127076	5425.76829734	269.76	0.0001	0.007898	82.7550
ERROR	372752	7497228.03349440	20.11317990		ROOT MSE		READ MEAN
CORRECTED TOTAL	372763	7556911.48476516			4.48477200		5.41933502

TABLE L5 SASSKGL4 18:35 6/5

GLM LINEAR REGRESSION OF MEAN NUMBER OF READING SKILLS ON GRADE LEVEL
(MATH SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: R SKILL MEAN # OF READING SKILLS PER GRADE
WEIGHT: NSTU

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	54968.39927958	54968.39927958	116.58	0.0001	0.920999	418.2618
ERROR	10	4715.05199120	471.50519912		ROOT MSE		R SKILL MEAN
CORRECTED TOTAL	11	59683.45127078			21.71417047		5.19152591

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.08757836	85.28	0.0001	0.07138121
GRADE	-0.15620710	-10.80	0.0001	0.01446750

TABLE L6 SASSKGL3 8:46 5/12

GLM POLYNOMIAL REGRESSION OF MEAN NUMBER OF READING SKILLS ON GRADE LEVEL
(READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: READ M
WEIGHT: READ N

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	58736.19946888	29368.09973444	279.03	0.0001	0.984129	197.6134
ERROR	9	947.25180190	105.25020021		ROOT MSE		READ M MEAN
CORRECTED TOTAL	11	59683.45127078			10.25915202		5.19152591

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	54968.39927958	522.26	0.0001	1	15440.85729539	146.71	0.0001
GRADE*GRADE	1	3767.80018930	35.80	0.0002	1	3767.80018930	35.80	0.0002

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.34035758	117.79	0.0001	0.05405823
GRADE	-0.29721284	-12.11	0.0001	0.02453525
GRADE*GRADE	0.01439507	5.98	0.0002	0.00240592

64

TABLE L7 SASTSG12 8:18 7/7

GLM ANOVA OF NUMBER OF READING SKILLS X GRADE LEVEL
(MATH SKILLS OR READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: READ		# OF READING SKILLS					
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	84286.07430629	7662.37039148	371.61	0.0001	0.009955	91.3859
ERROR	406547	8382653.57642584	20.61915000		ROOT MSE		READ MEAN
CORRECTED TOTAL	406558	8466939.65073212			4.54083142		4.96885569

TABLE L8 SASSKGL4 19:23 6/5

GLM LINEAR REGRESSION OF MEAN NUMBER OF READING SKILLS ON GRADE LEVEL
(MATH SKILLS OR READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: R SKILL WEIGHT: NSTU		MEAN # OF READING SKILLS					
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	79453.65652554	79453.65652554	164.42	0.0001	0.942666	469.6006
ERROR	10	4832.41778076	483.24177808		ROOT MSE		R SKILL MEAN
CORRECTED TOTAL	11	84286.07430630			21.98276093		4.68116127

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	5.74014337	82.79	0.0001	0.06933064
GRADE	-0.17836342	-12.82	0.0001	0.01391013

TABLE L9 SASSKGL1 14:49 5/12

GLM POLYNOMIAL REGRESSION OF MEAN NUMBER OF READING SKILLS ON GRADE LEVEL
(MATH SKILLS OR READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: R SKILL
WEIGHT: NSTU

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	81501.47782654	40750.73891327	131.71	0.0001	0.966963	375.7563
ERROR	9	2784.59647977	309.39960886		ROOT MSE		R MEAN MEAN
CORRECTED TOTAL	11	84286.07430630			17.58975864		4.68116127

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	79453.65652554	256.80	0.0001	1	14764.31276560	47.72	0.0001
GRADE*GRADE	1	2047.82130100	6.62	0.0301	1	2047.82130100	6.62	0.0301

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	5.51968202	66.40	0.0001	0.08314998
GRADE	-0.27726557	-6.91	0.0001	0.04011042
GRADE*GRADE	0.01007644	2.57	0.0301	0.00391671

262

TABLE L10 SASSKGL8 13:15 7/24

65

263

GLM ANOVA OF NUMBER OF READING SKILLS X GRADE LEVEL
(MATH SKILLS AND READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: READ C # OF READ CURRENT SKILLS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	11	52966.21018869	4815.11001715	214.55	0.0001	0.007712	83.1735
ERROR	303671	6815275.51433326	22.44295805		ROOT MSE		READ C MEAN
CORRECTED TOTAL	303682	6868241.72452195			4.73739993		5.69580451

TABLE L11 SASSKGL7 10:56 7/24

GLM LINEAR REGRESSION OF MEAN NUMBER OF READING SKILLS ON GRADE LEVEL
(MATH SKILLS AND READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: [illegible] MEAN
WEIGHT: [illegible] MSTU MEAN # OF READ SKILLS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	44828.63774364	44828.63774364	55.09	0.0001	0.846363	511.4158
ERROR	10	8137.57244507	813.75724451		ROOT MSE		R MEAN MEAN
CORRECTED TOTAL	11	52966.21018871			28.52643063		5.57793288

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.37786957	68.27	0.0001	0.10547254
GRADE	-0.16563621	-7.42	0.0001	0.02231644

TABLE L12 SASSKGL8 13:15 7/24

GLM POLYNOMIAL REGRESSION OF MEAN NUMBER OF READING SKILLS ON GRADE LEVEL
(MATH SKILLS AND READING SKILLS PRESENT IN EACH RECORD OF SKILLS DATA SET)

DEPENDENT VARIABLE: [illegible] MEAN
WEIGHT: [illegible] MSTU MEAN # OF READ SKILLS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	52302.24764715	26151.12382357	354.48	0.0001	0.987464	153.9846
ERROR	9	663.96254156	73.77361573		ROOT MSE		R MEAN MEAN
CORRECTED TOTAL	11	52966.21018871			8.58915687		5.57793288

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	44828.63774364	607.65	0.0001	1	19821.76378859	268.68	0.0001
GRADE*GRADE	1	7473.60990351	101.30	0.0001	1	7473.60990351	101.30	0.0001

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT	6.78715605	131.55	0.0001	0.05159556
GRADE	-0.48445671	-15.55	0.0001	0.03117005
GRADE*GRADE	0.02577706	10.07	0.0001	0.00256105

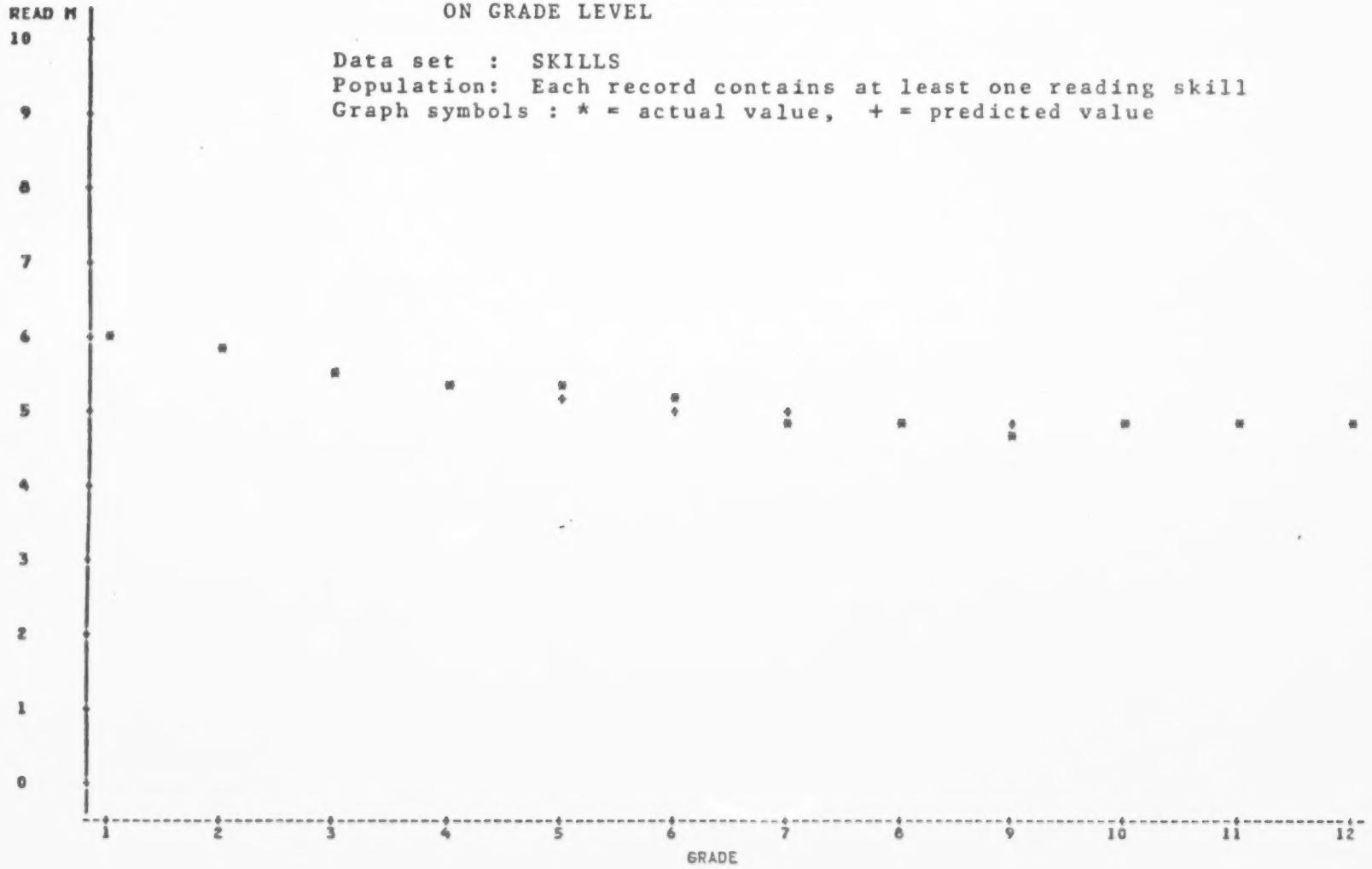
99

TABLE L13 SASSKGL8 13:15 7/24

SASTSG12

POLYNOMIAL REGRESSION OF MEAN NUMBER
OF READING SKILLS PER STUDENT
ON GRADE LEVEL

Data set : SKILLS
Population: Each record contains at least one reading skill
Graph symbols : * = actual value, + = predicted value

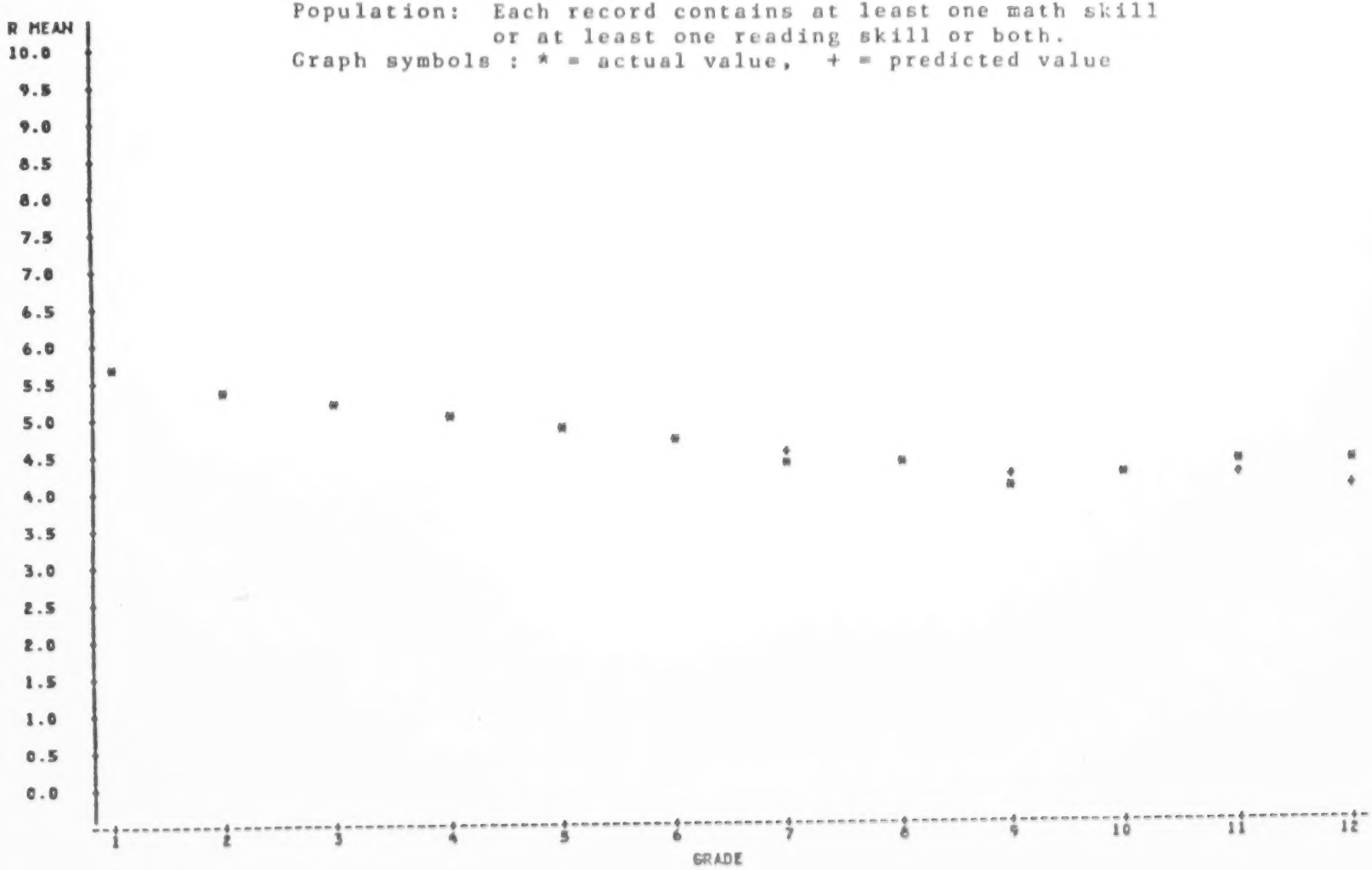


PLOT L1

67

POLYNOMIAL REGRESSION OF MEAN NUMBER
OF READING SKILLS PER STUDENT
ON GRADE LEVEL

Data set : SKILLS
Population: Each record contains at least one math skill
or at least one reading skill or both.
Graph symbols : * = actual value, + = predicted value

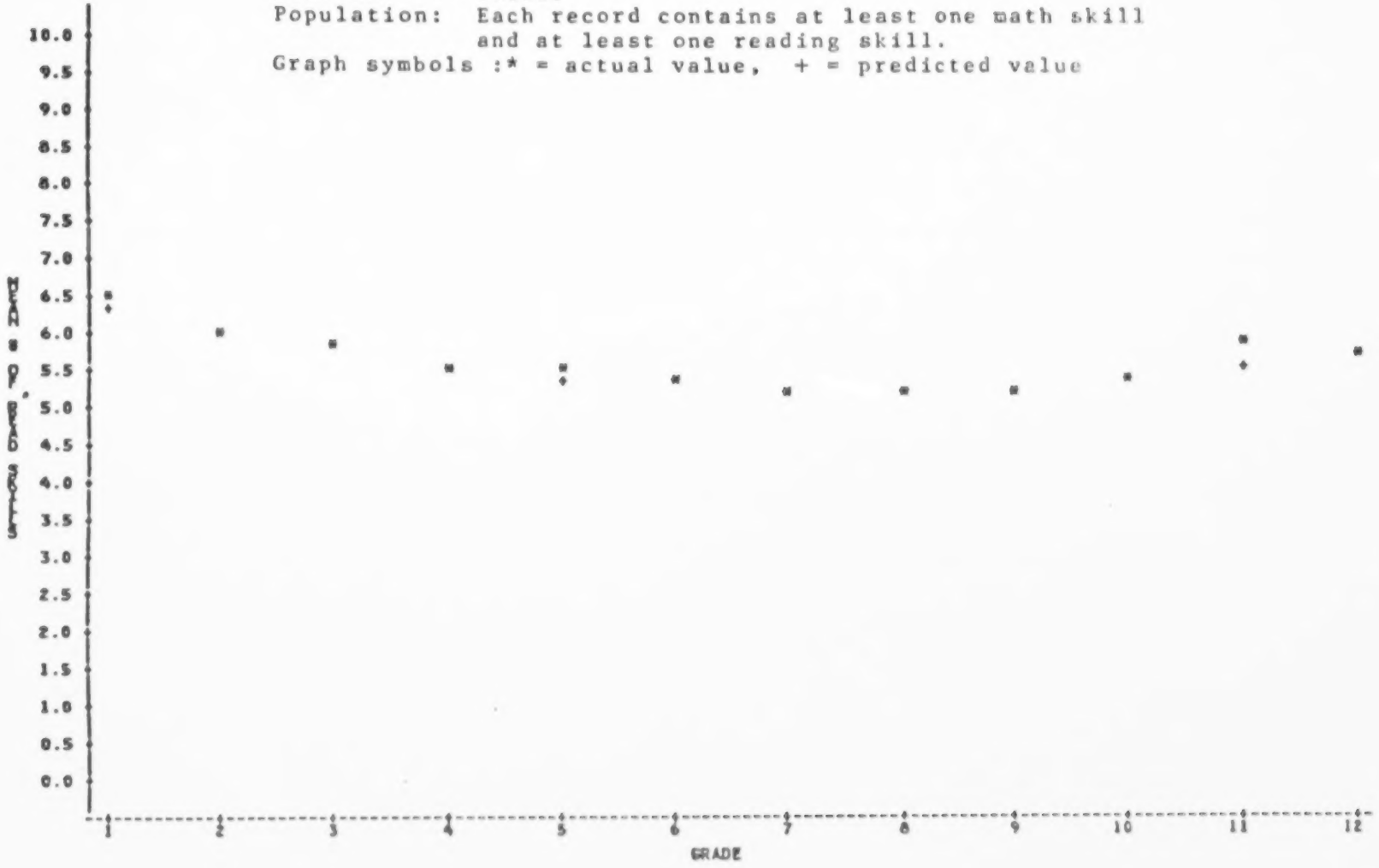


PLOT L2

POLYNOMIAL REGRESSION OF MEAN NUMBER
OF READING SKILLS PER STUDENT
ON GRADE LEVEL

13:15 THURSDAY, JULY 24, 1986 9
SASSKGL8

Data set : SKILLS
Population: Each record contains at least one math skill
and at least one reading skill.
Graph symbols :* = actual value, + = predicted value



PLOT L3

69

TABLE OF GRADE BY READ LEV

GRADE / READ LEV	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1	174329 12.03 91.70 63.18	12389 0.92 6.52 5.26	1951 0.14 1.03 0.86	591 0.04 0.31 0.31	415 0.03 0.22 0.27	218 0.02 0.11 0.19	107 0.01 0.06 0.15	39 0.00 0.02 0.08	11 0.00 0.01 0.06	20 0.00 0.01 0.18	34 0.00 0.02 0.51	5 0.00 0.00 0.15	190109 14.10
2	58703 4.35 29.13 21.29	127339 9.44 63.17 54.10	12013 0.89 5.96 5.32	1736 0.13 0.86 0.91	737 0.05 0.37 0.48	506 0.04 0.25 0.45	312 0.02 0.15 0.44	142 0.01 0.07 0.51	38 0.00 0.02 0.22	17 0.00 0.01 0.16	21 0.00 0.01 0.32	26 0.00 0.01 0.78	201590 14.95
3	21052 1.56 10.50 7.63	49262 3.65 25.05 20.93	112700 8.36 57.50 49.94	9859 0.73 5.01 5.17	1809 0.13 0.92 1.19	698 0.05 0.36 0.62	506 0.04 0.26 0.71	453 0.03 0.23 0.99	239 0.02 0.12 1.36	45 0.00 0.02 0.41	28 0.00 0.01 0.42	5 0.00 0.00 0.15	196656 14.59
4	9381 0.70 5.02 3.40	22435 1.66 12.02 9.53	47982 3.56 25.70 21.26	94840 7.03 50.80 49.70	8616 0.64 4.61 5.66	1656 0.12 0.89 1.46	400 0.03 0.21 0.56	498 0.04 0.27 1.08	353 0.03 0.19 2.01	410 0.03 0.22 3.77	107 0.01 0.06 1.61	20 0.00 0.01 0.60	186698 13.85
5	5580 0.41 3.21 2.02	11245 0.83 6.47 4.78	24719 1.83 14.22 10.95	39735 2.95 22.86 20.82	81727 6.06 47.02 53.69	8046 0.60 4.63 7.12	1012 0.08 0.58 1.43	476 0.04 0.27 1.04	314 0.02 0.18 1.79	477 0.04 0.27 4.39	390 0.03 0.22 5.86	88 0.01 0.05 2.63	173809 12.89
6	3491 0.26 2.35 1.27	6047 0.45 4.07 2.57	12748 0.95 8.58 5.65	21781 1.62 14.67 11.41	30386 2.25 20.46 19.96	67407 5.00 45.39 59.61	4767 0.35 3.21 6.73	895 0.07 0.60 1.95	283 0.02 0.19 1.61	277 0.02 0.18 2.55	234 0.02 0.16 3.52	191 0.01 0.13 5.70	148507 11.01
7	1863 0.14 1.71 0.68	3692 0.27 3.40 1.57	7314 0.54 6.73 3.24	11997 0.89 11.03 6.29	14894 1.10 13.70 9.78	18281 1.36 16.81 16.17	45398 3.37 41.76 64.07	4194 0.31 3.86 9.13	653 0.05 0.60 3.72	268 0.02 0.25 2.46	93 0.01 0.09 1.40	75 0.01 0.07 2.24	108722 8.06
8	1063 0.08 1.25 0.39	1967 0.15 2.31 0.84	4099 0.30 4.82 1.82	7039 0.52 8.28 3.69	9193 0.68 10.82 6.04	11846 0.88 13.94 10.48	13331 0.99 15.69 18.81	33787 2.51 39.76 73.55	2008 0.15 2.36 11.45	438 0.03 0.52 4.03	120 0.01 0.14 1.80	93 0.01 0.11 2.78	84984 6.30
9	282 0.02 0.36 0.10	607 0.05 0.77 0.26	1304 0.10 1.65 0.58	2018 0.15 2.55 1.06	2806 0.21 3.55 1.84	2644 0.20 3.34 2.34	3254 0.24 4.11 4.59	3290 0.24 4.16 7.16	10633 0.79 13.45 60.63	836 0.06 1.06 7.69	195 0.01 0.25 2.93	51 0.00 0.06 1.52	27920 2.07
10	109 0.01 0.70 0.04	279 0.02 1.80 0.12	492 0.04 3.17 0.22	770 0.06 4.96 0.40	1035 0.08 6.66 0.68	1138 0.08 7.32 1.01	1009 0.07 6.49 1.42	1319 0.10 8.49 2.87	1874 0.14 12.06 10.69	6692 0.50 43.07 61.52	672 0.05 4.33 10.10	152 0.01 0.98 4.54	15536 1.15
11	64 0.00 0.70 0.02	89 0.01 0.97 0.04	268 0.02 2.93 0.12	336 0.02 3.67 0.18	441 0.03 4.81 0.29	440 0.03 4.80 0.39	551 0.04 6.01 0.78	605 0.04 6.60 1.32	802 0.06 8.75 4.57	981 0.07 10.71 9.02	4175 0.31 45.57 62.73	410 0.03 4.48 12.24	9162 0.68
12	17 0.00 0.37 0.01	39 0.00 0.84 0.02	62 0.00 1.33 0.03	133 0.01 2.86 0.07	175 0.01 3.77 0.11	198 0.01 4.26 0.18	216 0.02 4.65 0.30	238 0.02 5.13 0.52	330 0.02 7.10 1.88	416 0.03 8.96 3.82	586 0.04 12.62 8.51	2235 0.17 48.12 66.70	4645 0.34
TOTAL	275934 20.46	235390 17.46	225652 16.74	190535 14.15	152234 11.29	113078 8.39	70858 5.26	45936 3.41	17538 1.30	10877 0.81	6655 0.49	3351 0.25	1348338 100.00

REGRESSION OF READING SKILL LEVEL ON GRADE SASSVFR2 10:18 TUESDAY, MAY 6, 1986 4

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: READ LEV
WEIGHT: COUNT FREQUENCY COUNT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	4736426.55799943	4736426.55799943	326.65	0.0001	0.697001	1852.5585
ERROR	142	2059012.18137675	14500.08578434		ROOT MSE		READ LEV MEAN
CORRECTED TOTAL	143	6795438.73937618			120.41630199		6.50000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	4736426.55799943	326.65	0.0001	1	4736426.55799943	326.65	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	0.31186178	1.49	0.1397	0.20998826
GRADE	0.78143675	18.07	0.0001	0.04323682

TABLE M2

POLYNOMIAL REGRESS OF READING SKILL LEVEL ON GRADE SASSVFR2 10:18 TUESDAY, MAY 6, 1986. 2

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: READ LEV
WEIGHT: COUNT FREQUENCY COUNT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	4736430.35912973	2368215.17956486	162.17	0.0001	0.697001	1859.1145
ERROR	141	2059008.38024645	14602.89631380		ROOT MSE		READ LEV MEAN
CORRECTED TOTAL	143	6795438.73937618			120.84244417		6.50000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
GRADE	1	4736426.55799943	324.35	0.0001	1	366231.21293183	25.08	0.0001
GRADE*GRADE	1	3.80113030	0.00	0.9872	1	3.80113030	0.00	0.9872

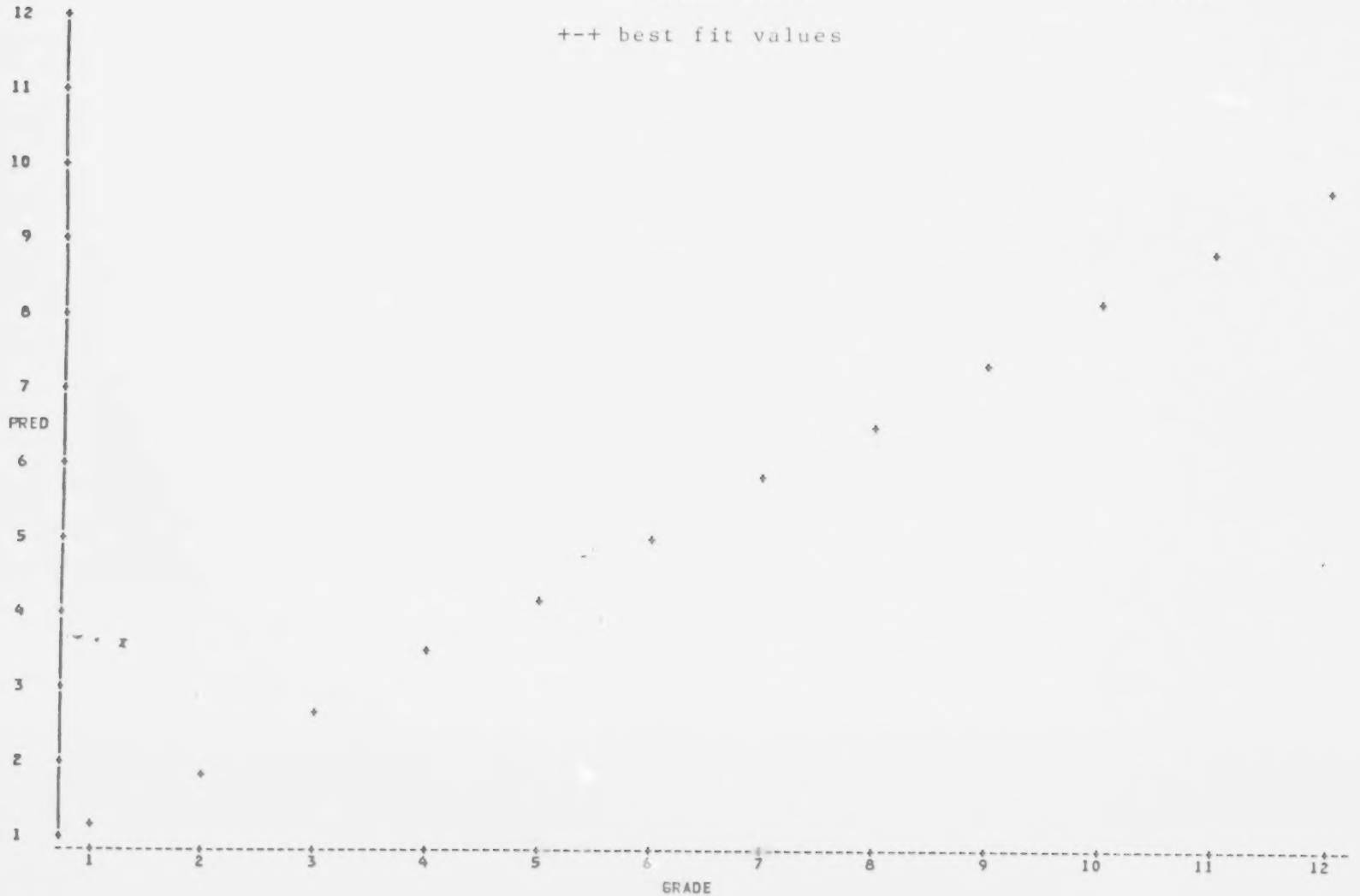
PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	0.31619718	0.93	0.3561	0.34149017
GRADE	0.77902661	5.01	0.0001	0.15553663
GRADE*GRADE	0.00024772	0.02	0.9872	0.01535379

TABLE M3

SIMPLE REGRESSION OF READING LEVEL ON GRADE LEVEL

SASSVTR2

+-+ best fit values



276

PLOT M1

277

SUMMARY PARAMETERS OF THE SKILLS DATA SET AND THE SKILLS/TEST DATA SET

From the MSRTS data base the below data sets were created. The SKILLS data set contains 406,559 grade 1-12 skill records. Each record contains at least one math skill or reading skill or both. The TEST Data Set contains 351,011 records each of which contains data on at least one math test or one reading test or both.

READING SKILL CONTENT IN THE SKILLS Data Set		READING TEST CONTENT IN THE TEST FILE	
number of records /w reading skills	number of reading skills	number of records in IDSHTD	number of reading tests
372,764	2,020,133	351,011	211,331

The above two data sets (including the reading skills and reading tests) were integrated into the combined SKILL/ TEST data set whose Reading Skill Test content is shown below.

READING SKILL AND READING TEST CONTENT OF THE SKILLS/TEST DATA SET (IDSHTDSK)

record contents	number of records	number of skills	number of tests
tests with no skills	62,778	0	93,949
skills with no matching tests	1,844	11,021	0
skills with matching tests	76,515	446,226	117,382
total	141,137	457,247	211,331
total reading tests in file	139,293		211,331
total reading skills in file	78,359	457,247	

Compiled from TSM11 7:25 6/6

TABLE N1

ANALYSIS SUMMARY TABLE

Data Set : Skills/Test (IDSHTDSK) 74
Cardinality: 144,148
Model : MEAN NUMBER OF READING SKILLS = GRADE LEVEL

TYPE OF ANALYSIS-> ANOVA			REGRESSION				
			LINEAR		POLYNOMIAL		
TERM -> ->	P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
1 READ_C	S	.0240	-.29	.9653	-.2834	N	.9680
2 MATH_C OR READ_C	S	.0283	-.32	.9619	-.3266	N	.9619
3 MATH_C AND READ_C	S	.0242	-.32	.9463	-.3919	N	.9491
4 READ_T	S	.0766	-.41	.9218	N	-.0272	.9589
5 MATH_T OR READ_T	S	.0764	-.41	.9248	N	-.0262	.9596
6 MATH_T AND READ_T	S	.0725	-.41	.9278	N	-.0258	.9613
7 MATH_C AND MATH_T	S	.0233	-.28	.9645	-.3384	N	.9668
8 MATH_C OR READ_C AND READ_T	S	.0277	-.31	.9607	-.3137	N	.9607

S = significant N = not significant +,- = direction of (significant) term

TABLE N2

****

ANALYSIS SUMMARY TABLE

Data Set : Skills/Test (IDSHTDSK)
Cardinality: 144,148
Model : MEAN NUMBER OF READING TESTS = GRADE LEVEL

TYPE OF ANALYSIS-> ANOVA			REGRESSION				
			LINEAR		POLYNOMIAL		
TERM -> ->	P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
1 READ_C	S	.0068	N	.0462	+.0842	-.0081	.8647
2 MATH_C OR READ_C	S	.0062	N	.0113	+.0792	-.0078	.8402
3 MATH_C AND READ_C	S	.0066	N	.2142	+.0904	-.0085	.8578
4 READ_T	S	.0078	N	.2118	+.0510	-.0052	.8747
5 MATH_T OR READ_T	S	.0083	N	0944	+.0641	-.0062	.8497
6 MATH_T AND READ_T	S	.0095	N	.2926	+0534	-.0057	.9106
7 READ_C AND READ_T	S	.0050	N	.0002	+.0662	-.0067	.8493
8 MATH_C OR READ_C AND READ_T	S	.0049	N	.0014	+.0641	-.0066	.8495

S = significant N = not significant +,- = direction of (significant) term

TABLE N3

ANALYSIS SUMMARY TABLE

Data Set : Skills/Test (IDSHTDSK)
Cardinality: 144,148
Model : MEAN NUMBER OF READING TESTS =.READING SKILLS PER STUDENT

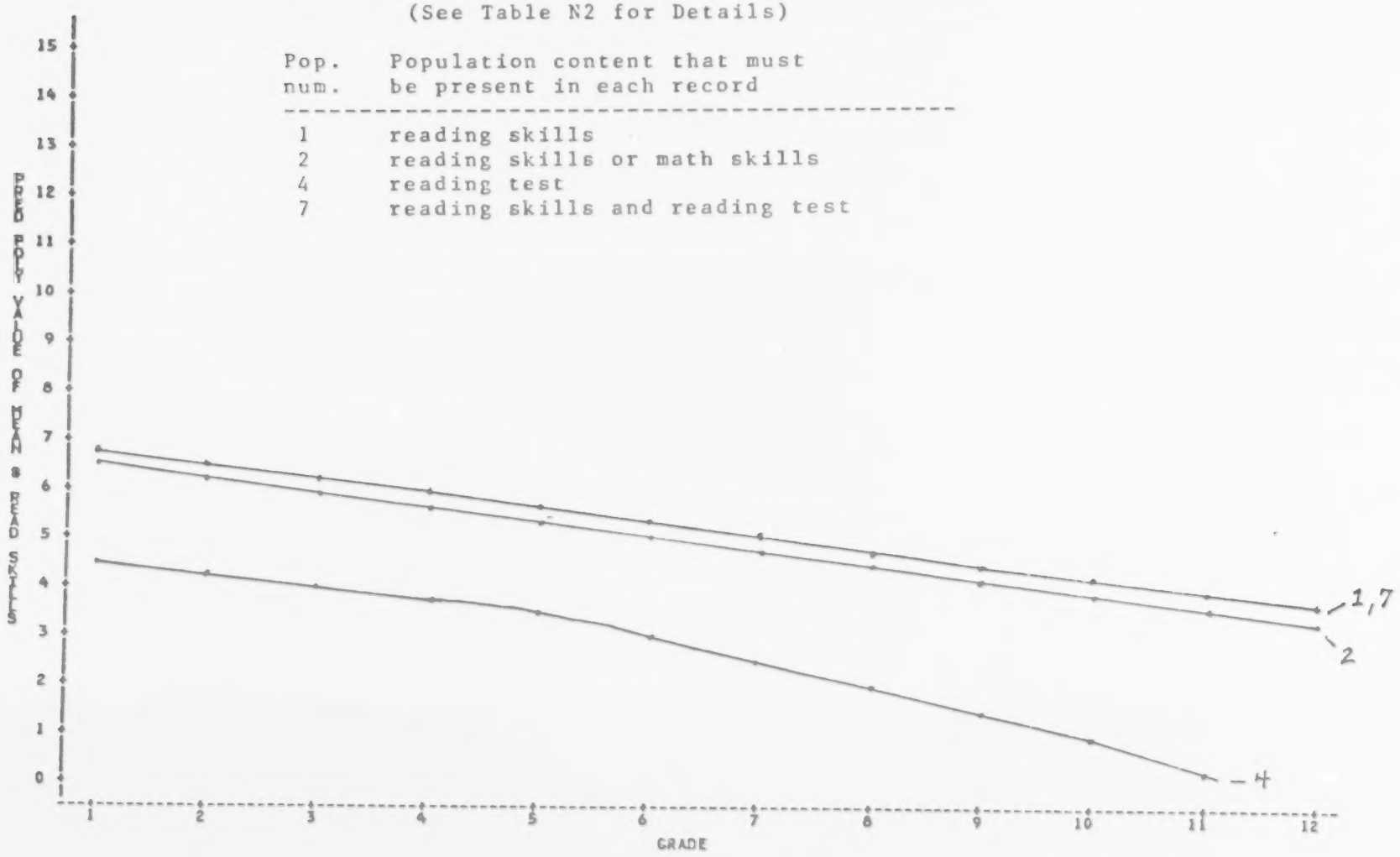
TYPE OF ANALYSIS->	ANOVA	REGRESSION					
			LINEAR		POLYNOMIAL		
TERM -> ->	P>F	R-SQ	SLOPE	R-SQ	LINEAR	QUADRATIC	R-SQ
POPULATIONS:							
1 READ_C	S	.0099	N	.0070	N	N	.2901
2 MATH_C OR READ_C	S	.0257	N	.0807	+.0349	-.0013	.3875
3 MATH_C AND READ_C	S	.0098	N	.0003	N	N	.2564
4 READ_T	S	.0071	N	.0931	+.0149	-.0006	.3729
5 MATH_T OR READ_T	S	.0084	N	.1617	+.0202	-.0008	.5087
6 MATH_T AND READ_T	S	.0068	N	.1071	+.0160	-.0006	.4527
7 READ_C AND READ_T	S	.0101	N	.0144	N	N	.3124
8 MATH_C OR READ_C AND READ_T	S	.0105	N	.0338	+.0200	-.0008	.3602

S = significant N = not significant +,- = direction of (significant) term

TABLE N4

BEST FIT CURVE FOR THE REGRESSION OF THE MEAN
NUMBER OF READING SKILLS PER STUDENT
ON GRADE LEVEL
FOR FOUR POPULATIONS OF THE SKILLS/TEST DATA SET
(See Table N2 for Details)

Pop. num.	Population content that must be present in each record
1	reading skills
2	reading skills or math skills
4	reading test
7	reading skills and reading test



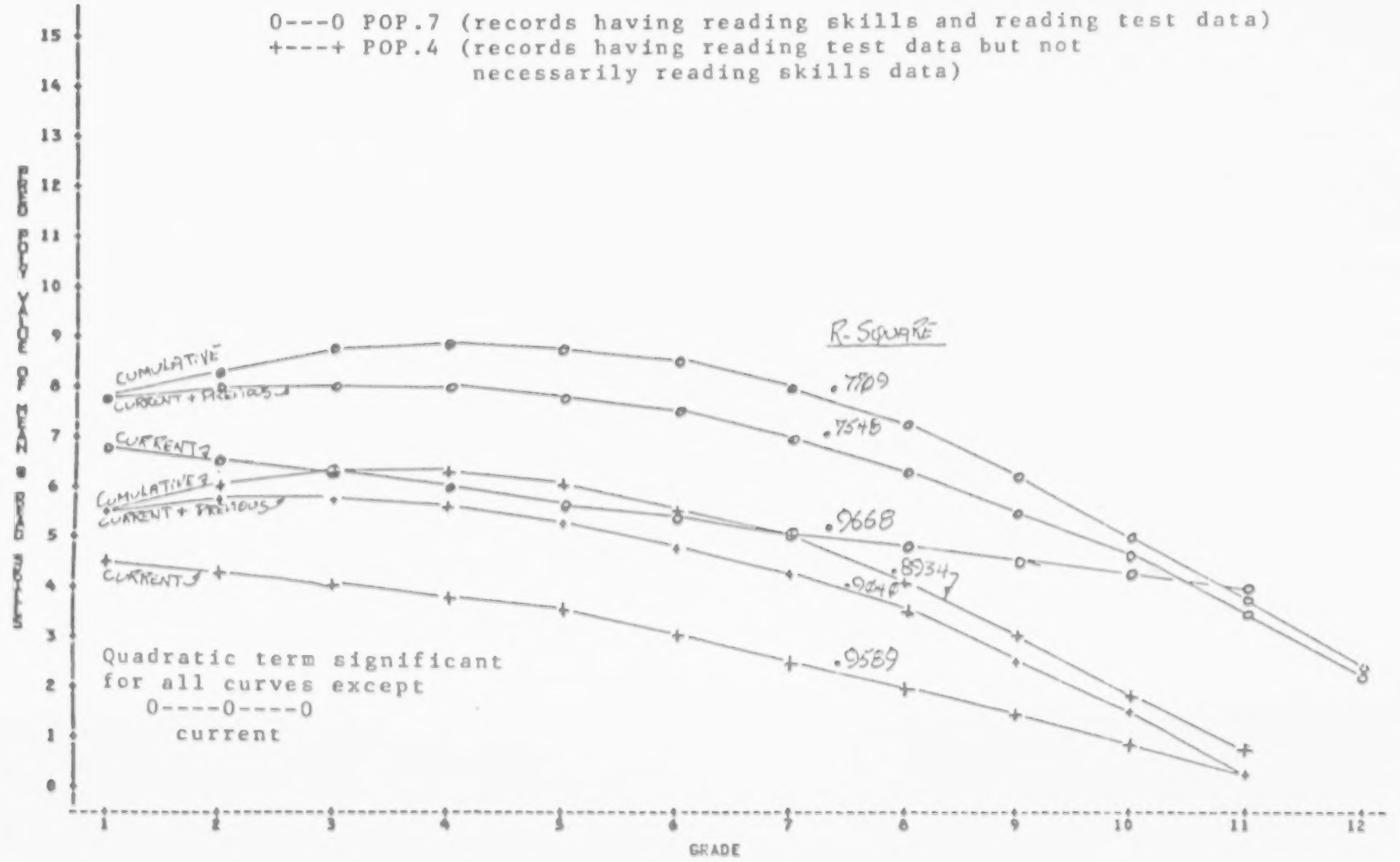
PLOT N1

76

BEST FIT POLYNOMIAL OF MEAN NUMBER OF
CURRENT, CURRENT + PREVIOUS AND CUMULATIVE
READING SKILLS PER STUDENT

13:39 FRIDAY, JUNE 6, 1986 20

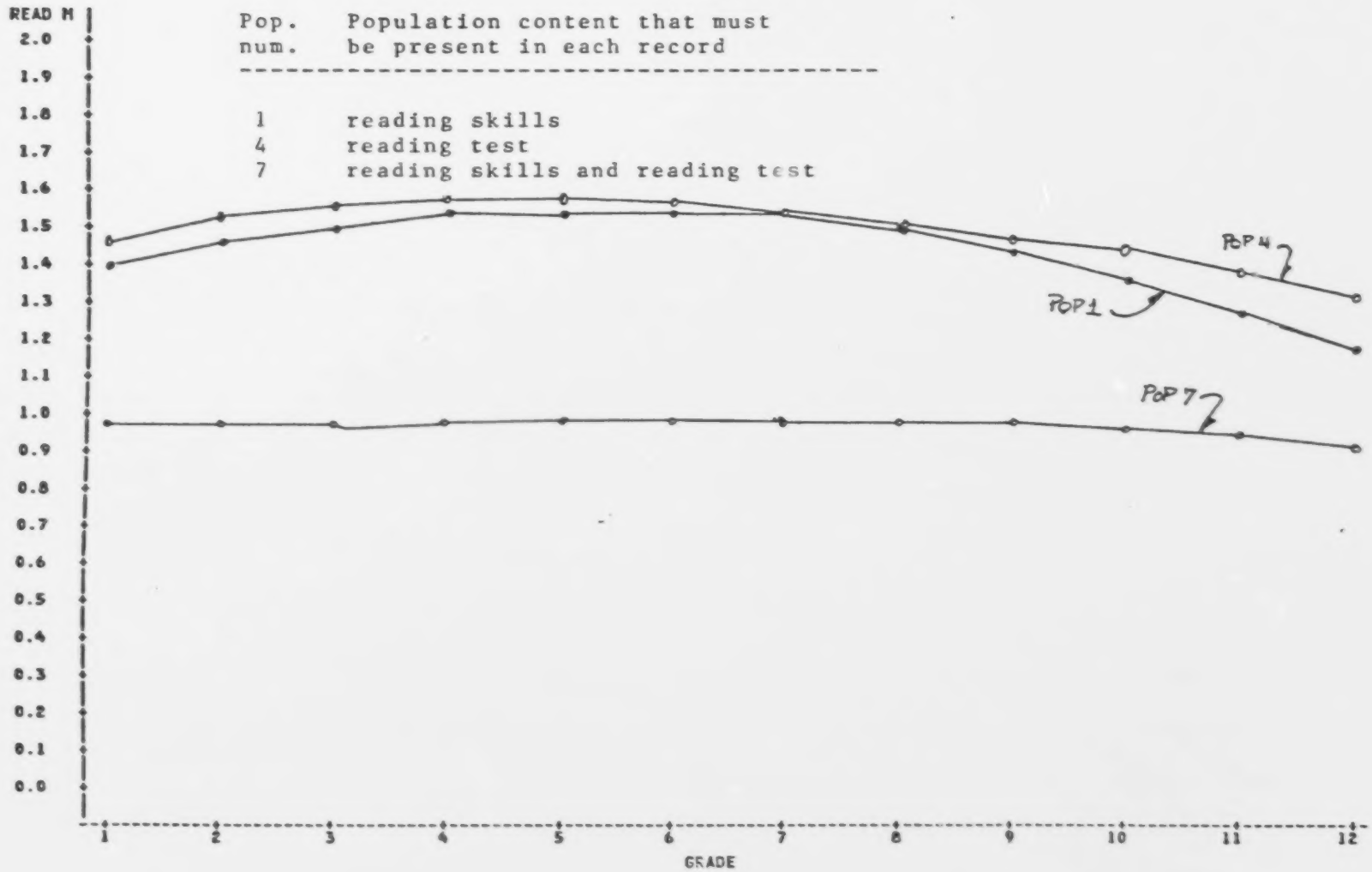
SASTSMN10.1



PLOT N2

BEST FIT CURVE FOR THE REGRESSION OF THE MEAN
NUMBER OF READING TESTS PER STUDENT
ON GRADE LEVEL
FOR THREE POPULATIONS OF THE SKILLS/TEST DATA SET
(See Table N3 for Details)

15:21 MONDAY, JULY 21, 1986 10
SASTSGL8

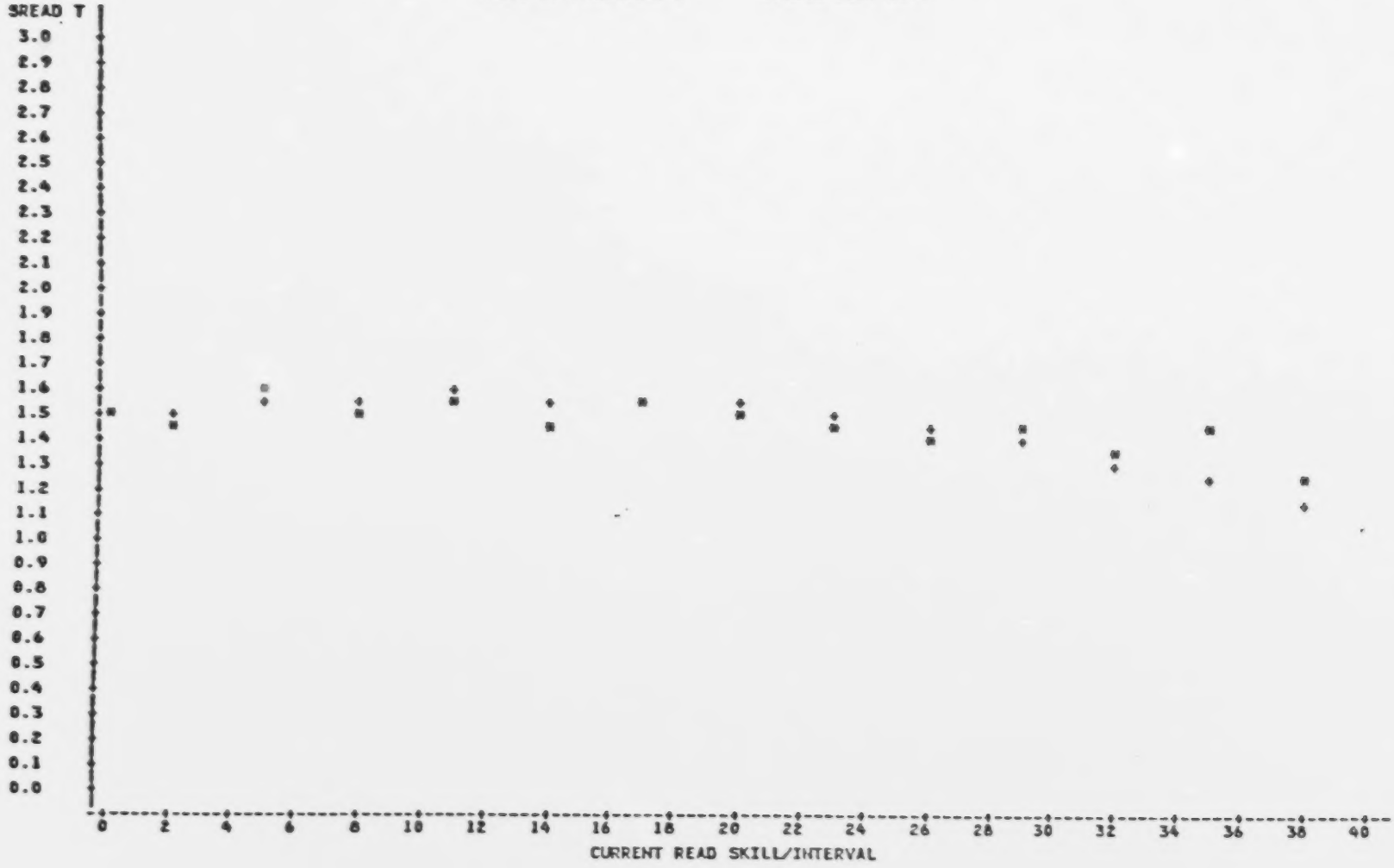


PLOT N3

POLYNOMIAL REGRESSION # OF READ TEST
ON CURRENT GRADE READING SKILL INTERVALS SASTSM23

IF READ TEST > 0

PLOT OF SREAD T*READ I SYMBOL USED IS *
PLOT OF PRED*READ I SYMBOL USED IS +



PLOT N4

SUMMARY TABLES OF THREE POPULATIONS OF SCORES ON READING TESTS
AS A FUNCTION OF THE NUMBER OF READING SKILLS IN STUDENTS RECORDS 80
IN THE SKILLS/TEST DATA SET

C U R R E N T READING S K I L L S

	Type of Regression Model					
	Linear		Polynomial			
	slope	R-sq.	linear term	quad. term	R-sq	
GRADE						
01	NS	.00	NS	NS	.01	
02	NS	.20	NS	NS	.22	
03	-0.36	.41	-0.63	NS	.50	
04	-0.43	.57	-0.88	+0.02	.80	POPULATION 1:
05	-0.57	.78	-0.71	NS	.80	Records containing at least one
06	-0.57	.71	-0.66	NS	.72	reading test but not necessarily
07	-0.76	.49	-1.43	+0.04	.68	reading skill data
08	-0.62	.38	-1.18	NS	.51	
09	-2.48	.66	-5.27	+0.40	.94	
10	NS	.17	NS	NS	.44	TABLE 01
11	NS	.49	NS	NS	.77	
12	NS	.12	NS	NS	.12	
GRADE						
01	NS	.05	NS	NS	.15	
02	-0.34	.43	NS	NS	.46	
03	-0.46	.45	-1.21	+0.03	.74	
04	-0.37	.39	-1.08	+0.03	.75	POPULATION 2:
05	-0.54	.68	-0.77	NS	.71	Records containing at least one
06	-0.56	.59	-0.71	NS	.60	reading skill and at least one
07	NS	.09	NS	NS	.11	reading test.
08	NS	.07	NS	NS	.13	
09	NS	.03	NS	NS	.74	
10	NS	.64	+6.29	NS	.95	TABLE 02
11	NS	.08	NS	NS	.10	
12	NS	.02	NS	NS	.65	
GRADE						
01	NS	.01	NS	NS	.05	
02	-0.31	.39	NS	NS	.40	
03	-0.52	.47	-1.32	+0.03	.77	
04	-0.39	.42	-1.03	+0.03	.76	POPULATION 3:
05	-0.57	.70	-0.82	NS	.74	Records containing at least one
06	-0.50	.52	NS	NS	.52	math skill or one reading skill
07	NS	.18	NS	NS	.27	or both and one reading test.
08	NS	.05	NS	NS	.07	
09	NS	.23	-2.79	+0.22	.87	
10	NS	.36	NS	NS	.36	TABLE 03
11	NS	.07	NS	NS	.10	
12	NS	.02	NS	NS	.45	

SUMMARY TABLES OF ONE POPULATION OF SCORES ON READING TESTS AS A FUNCTION OF THE NUMBER OF READING SKILLS IN STUDENTS RECORDS IN THE SKILLS/TEST DATA SET

CURRENT + PREVIOUS READING SKILLS

	Type of Regression Model				
	Linear		Polynomial		
	slope	R-sq.	linear term	quad. term	R-sq
GRADE					
01	NS	.00	NS	NS	.00
02	NS	.07	NS	NS	.08
03	-0.24	.41	-0.45	NS	.49
04	NS	.18	-0.54	NS	.42
05	-0.36	.67	-0.471	NS	.69
06	-0.25	.43	NS	NS	.43
07	-0.37	.30	NS	NS	.38
08	-0.35	.34	-0.88	+0.03	.60
09	-1.28	.65	-2.57	+0.13	.90
10	NS	.21	NS	NS	.44
11	NS	.23	NS	NS	.55
12	NS	.08	NS	NS	.09

POPULATION 1:
Records containing at least one reading test but not necessarily reading skill data.

TABLE 04

C U M U L A T I V E READING S K I L L S

	Type of Regression Model				
	Linear		Polynomial		
	slope	R-sq.	linear term	quad. term	R-sq
GRADE					
01	NS	.00	NS	NS	.01
02	NS	.07	NS	NS	.08
03	-0.21	.40	-0.42	NS	.48
04	NS	.27	-0.48	NS	.48
05	-0.33	.79	-0.41	NS	.80
06	-0.20	.45	NS	NS	.46
07	NS	.27	NS	NS	.36
08	-0.35	.37	-0.71	NS	.49
09	-0.95	.56	-1.80	+0.06	.77
10	NS	.22	-0.94	NS	.35
11	NS	.13	NS	NS	.51
12	NS	.00	NS	NS	.16

POPULATION 1:
Records having at least one reading test but not necessarily reading skill data.

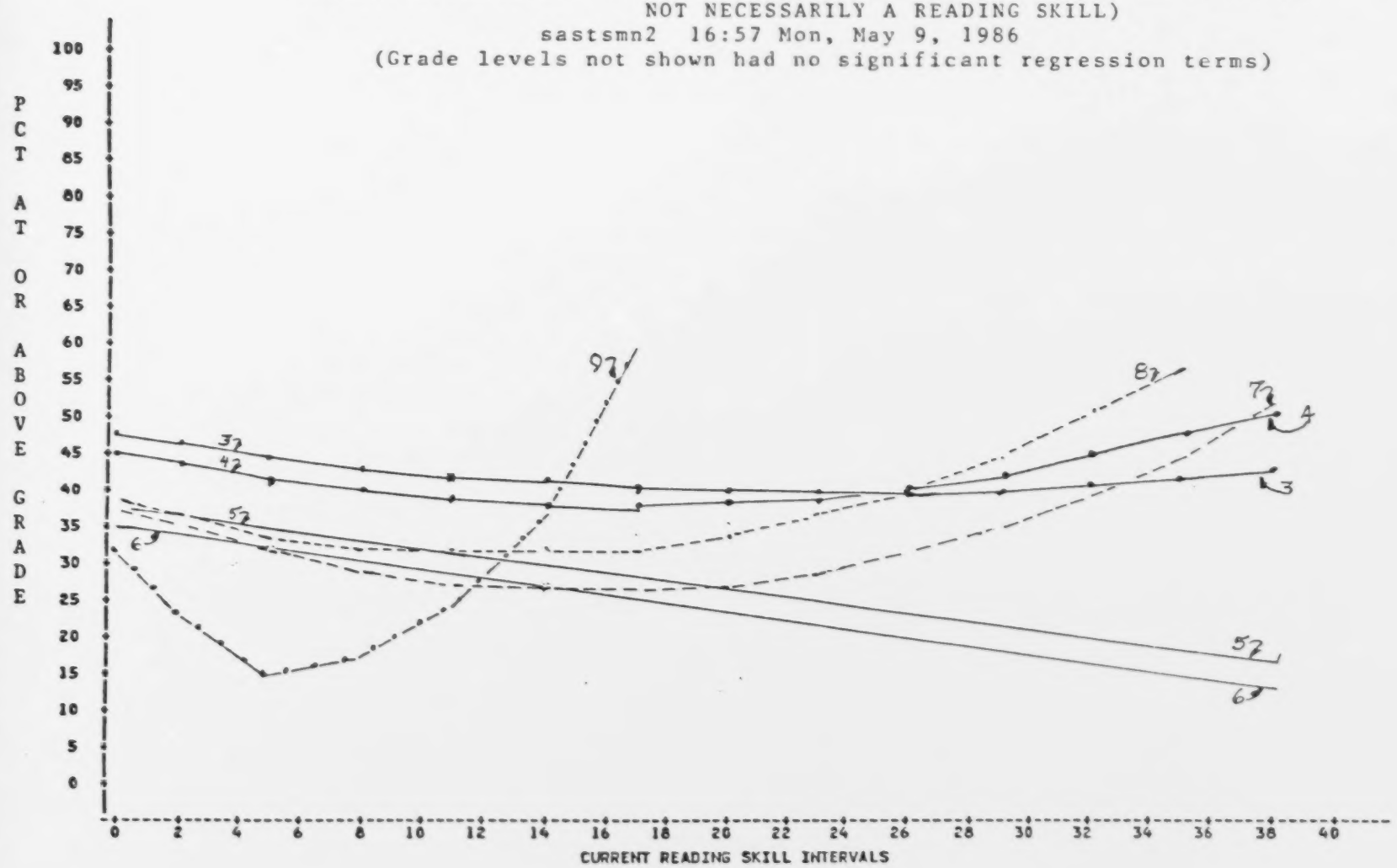
TABLE 05

NUMBER OF CASES FOR EACH READING SKILL INTERVAL BY GRADE FOR POPULATION 1

G.L.	0	2	5	8	11	14	17	20	23	26	29	32	35	38
01	4549	3213	3166	926	1100	309	228	99	92	43	33	20	15	25
02	5634	3828	3771	1113	1196	299	178	98	58	37	31	11	13	29
03	5066	4310	3828	1131	870	308	177	82	45	25	20	8	12	38
04	6096	4734	3833	1070	914	187	142	54	37	25	11	6	17	10
05	5835	4707	3701	954	778	180	117	48	44	14	20	5	8	10
06	5801	3906	3526	648	633	117	80	33	29	11	11	10	5	9
07	6536	3409	2424	443	373	77	50	33	14	8	2	3	2	5
08	5622	2371	2061	246	272	50	37	15	13	3	1	-	2	-
09	6148	1097	470	40	43	5	3	-	-	-	-	-	-	-
10	5242	572	243	20	21	2	-	-	-	-	-	-	-	-
11	3765	325	119	6	9	1	-	-	-	-	-	-	-	-
12	1711	139	47	1	7	1	1	-	-	-	-	-	-	-

TABLE 06

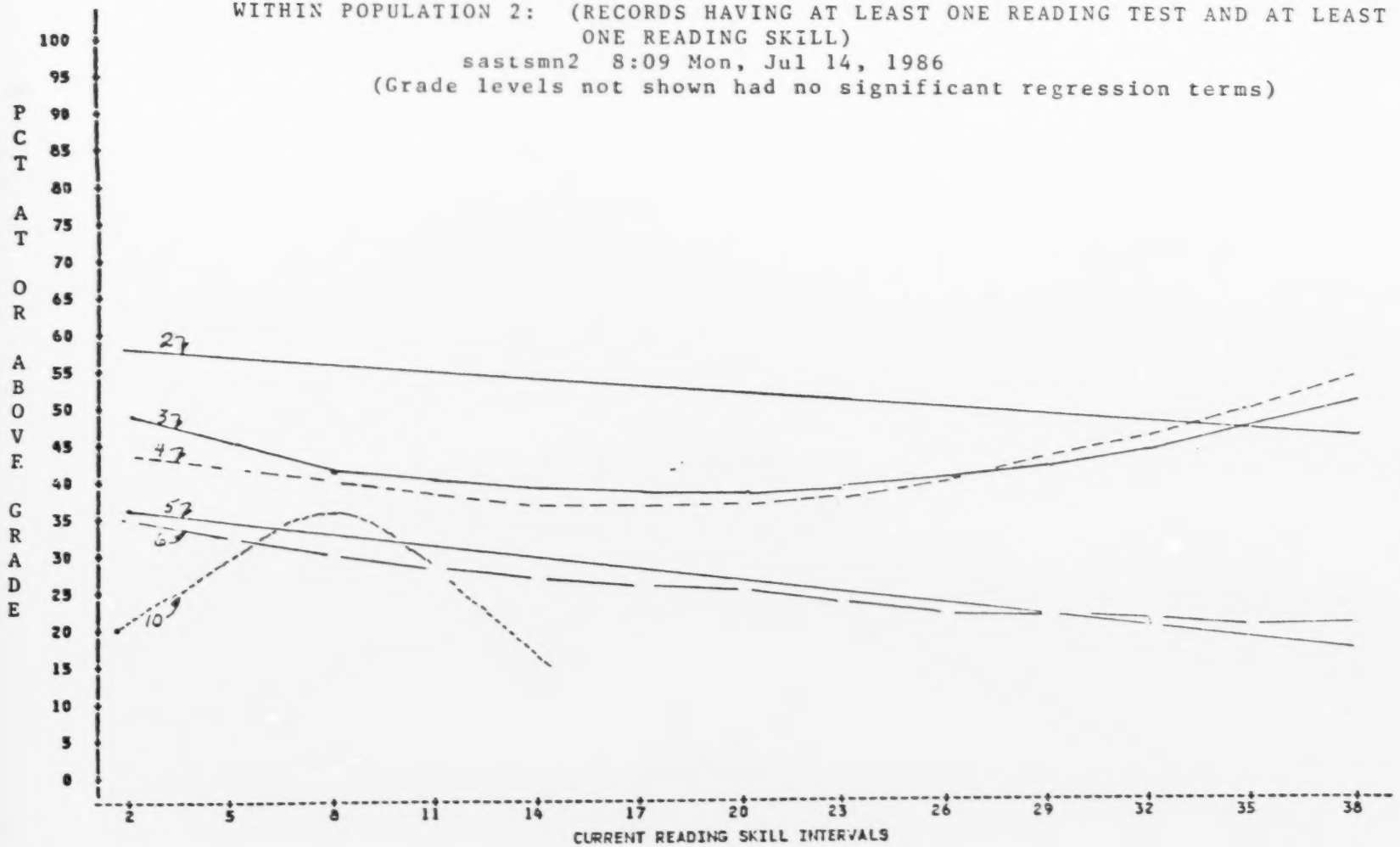
REGRESSION CURVES OF BEST FIT OR
PERCENT OF READING SCORES AT OR ABOVE GRADE LEVEL
ON THE
NUMBER OF CURRENT READING SKILLS IN STUDENT RECORDS FOR GRADE LEVELS
WITHIN POPULATION 1: (RECORDS HAVING AT LEAST ONE READING TEST BUT
NOT NECESSARILY A READING SKILL)
sastsmn2 16:57 Mon, May 9, 1986
(Grade levels not shown had no significant regression terms)



PLOT 01 See TABLE 01 for details

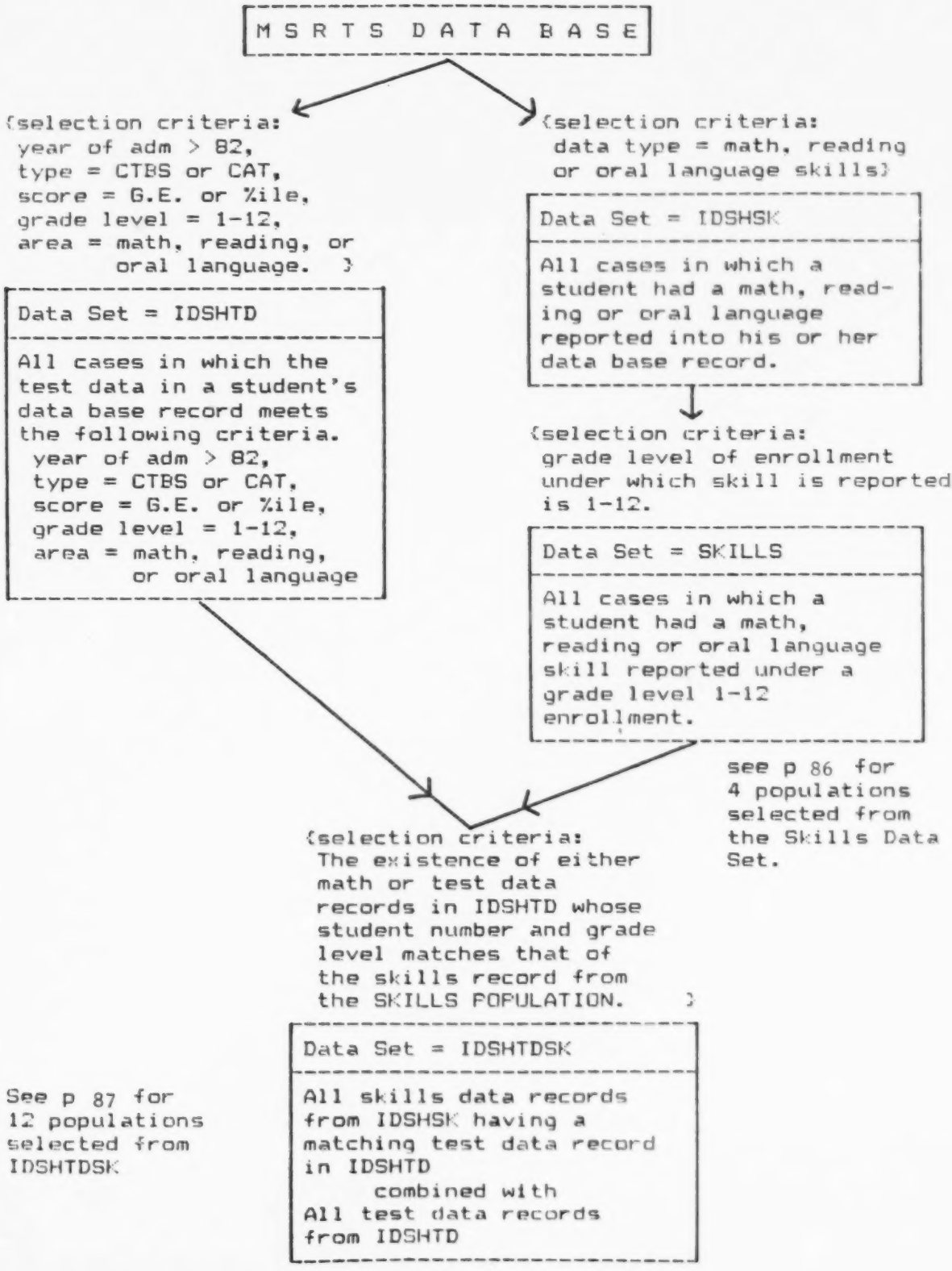
83

REGRESSION CURVES OF BEST FIT OR
PERCENT OF READING SCORES AT OR ABOVE GRADE LEVEL
ON THE
NUMBER OF CURRENT READING SKILLS IN STUDENT RECORDS FOR GRADE LEVELS
WITHIN POPULATION 2: (RECORDS HAVING AT LEAST ONE READING TEST AND AT LEAST ONE READING SKILL)
sastsmn2 8:09 Mon, Jul 14, 1986
(Grade levels not shown had no significant regression terms)



PLOT 02 See TABLE 02 for details

DATA SETS SELECTED FROM THE MSRTS DATA BASE
FOR THE ANALYSIS OF POPULATIONS OF
SKILLS AND TEST DATA


MSRTS DATA BASE
{selection criteria:
year of adm > 82,
type = CTBS or CAT,
score = G.E. or %ile,
grade level = 1-12,
area = math, reading, or
oral language. }
Data Set = IDSHTD
All cases in which the test data in a student's data base record meets the following criteria.
year of adm > 82,
type = CTBS or CAT,
score = G.E. or %ile,
grade level = 1-12,
area = math, reading, or oral language
{selection criteria:
data type = math, reading or oral language skills}
Data Set = IDSHSK
All cases in which a student had a math, reading or oral language reported into his or her data base record.
{selection criteria:
grade level of enrollment under which skill is reported is 1-12.
Data Set = SKILLS
All cases in which a student had a math, reading or oral language skill reported under a grade level 1-12 enrollment.
see p 86 for 4 populations selected from the Skills Data Set.
{selection criteria:
The existence of either math or test data records in IDSHTD whose student number and grade level matches that of the skills record from the SKILLS POPULATION. }
Data Set = IDSHTDSK
All skills data records from IDSHSK having a matching test data record in IDSHTD
combined with
All test data records from IDSHTD
See p 87 for 12 populations selected from IDSHTDSK

Appendix Figure C-1

4 POPULATIONS SELECTED FROM DATA SET SKILLS
by using (SELECTION CRITERIA)

(MATH_C)	(READ_C)
MATH SKILLS POPULATION	READING SKILLS POPULATION
All cases in which a student had math skills reported under a grade level 1-12 enrollment.	All cases in which a student had reading skills reported under a grade level 1-12 enrollment.

(MATH_C OR READ_C)	(MATH_C AND READ_C)
MATH OR READING SKILLS POPULATION	MATH AND READING SKILLS POPULATION
All cases in which a student had math or reading skills reported under a grade level 1-12 enrollment.	All cases in which a student had math and reading skills reported under a grade level 1-12 enrollment.

MATH_C = Select each record having at least one math skill
READ_C = Select each record having at least one reading skill

Appendix Figure C-2

12 POPULATIONS SELECTED FROM IDSHTDSK
by using (SELECTION CRITERIA)

(MATH_C)	(MATH_C OR READ_C)	(READ_C AND READ_T)
MATH SKILLS/TEST POPULATION	MATH OR READING SKILLS/ TEST POPULATION	READING SKILLS/READING TEST POPULATION
All cases in which a student had math skills reported under a grade level 1-12 enrollment, and had math or reading test data reported at the same grade level.	All cases in which a student had math or reading skills reported under a grade level 1-12 enrollment, and had math or reading test data reported at the same grade level.	All cases in which a student had reading skills reported under a grade level 1-12 enrollment, and had reading test data reported at the same grade level.
(READ_C)	(MATH_C AND READ_C)	(MATH_T AND READ_T)
READING SKILLS/TEST POPULATION	MATH AND READING SKILLS/ TEST POPULATION	SKILLS/MATH AND READING TEST POPULATION
All cases in which a student had reading skills reported under a grade level 1-12 enrollment, and had math or reading test data reported at the same grade level.	All cases in which a student had math and reading skills reported under a grade level 1-12 enrollment, and had math or reading test data reported at the same grade level.	All cases in which a student had math and reading test data reported under a grade 1-12 enrollment.
(MATH_T)	(MATH_C AND MATH_T)	(MATH_T OR READING_T)
SKILLS/MATH TEST POPULATION	MATH SKILLS/MATH TEST POPULATION	SKILLS/MATH OR READING TEST POPULATION
All cases in which a student had math test data reported under a grade level 1-12 enrollment.	All cases in which a student had math skills reported under a grade level 1-12 enrollment, and had math test data reported at the same grade level.	All cases in which a student had math or reading test data reported under a grade level 1-12 enrollment.
(READ_T)	((MATH_C OR READ_C) AND (READ_T))	((MATH_C OR READ_C) AND (MATH_C))
SKILLS/READING TEST POPULATION	MATH OR READING SKILLS/ READING TEST POPULATION	MATH OR READING SKILLS/ MATH TEST POPULATION
All cases in which a student had reading test data reported under a grade level 1-12 enrollment.	All cases in which a student had math or reading skills reported under a grade level 1-12 enrollment and had reading test data reported at the same grade level.	All cases in which a student had math or reading skills reported under a grade level 1-12 enrollment and had math test data reported at the same grade level.

Appendix Figure C-3

Number of Records making up the Populations
Selected from the SKILLS and IDSHTDSK data sets

POPULATION	num. of SKILLS records	%	num. of IDSHTDSK records	%
MATH_C	337,478	83.0	71,360	49.5
READ_C	372,764	91.7	78,359	54.4
MATH_C or READ_C	406,559	100.0	84,024	58.3
MATH_C AND READ_C	303,683	74.7	65,695	45.5
MATH_T			131,868	91.5
READ_T			139,293	96.6
MATH_C AND MATH_T			66,734	46.3
READ_C AND READ_T			76,515	53.1
MATH_T OR READ_T			144,148	100.0
MATH_T AND READ_T			127,013	88.1
MATH_C OR READ_C AND MATH_T			76,565	53.1
MATH_C OR READ_C AND READ_T			80,918	56.1

SKILLS has 406,559 records and IDSHTDSK has 144,148 records

Appendix Figure C-4

Compiled from TSM11 7:25 6/6

Number of records and skills in Populations
Selected from the SKILLS data set

population	number of records	number of math skills	number of reading skills
MATH_C	337,478	2,441,431	NA
READ_C	372,764	NA	2,020,133
MATH_C OR READ_C	406,559	2,441,431	2,020,133
MATH_C AND READ_C	303,683	2,252,761	1,729,719

Appendix Figure C-5

compiled from SKGL4 18:35 6/5 and SKGL8 13:15 7/24

Number of Records, Skills and Tests in Populations
Selected from the IDSHTDSK data set

population	number of records	number of math skills	number of math tests	number of reading skills	number
MATH_C	71,360	633,587		402,559	
READ_C	78,359	600,625		457,247	
MATH_C OR READ_C	84,024	633,587		457,247	
MATH_C AND READ_C	65,695	600,625		402,599	
MATH_T	131,868	602,216		419,431	
READ_T	139,293	610,338		446,226	
MATH_C AND MATH_T	66,734	602,216	72,043	376,752	99,593
READ_C AND READ_T	76,515	585,794	74,620	446,226	117,373
MATH_T OR READ_T	144,148	633,587	139,681	457,247	211,331
MATH_T AND READ_T	127,013	578,967	134,487	408,410	195,492
MATH_C OR READ_C AND MATH_T	76,565	602,216	82,339	414,431	114,252
MATH_C OR READ_C AND READ_T	80,918	610,338	78,956	446,226	123,775

Appendix Figure C-6

Compiled from TSM11 7:25 6/6

Number of Records making up the Mutually Exclusive Populations Selected from the SKILLS data sets

ID num	selection criteria: MATH_C	READ_C	num of records	percent of total	number of skills math	reading
1	>0	=0	33,795	8.31	188,670	0
2	=0	>0	69,081	16.99	0	290,414
3	>0	>0	303,683	74.70	2,252,761	1,729,719
		Totals	406,559	100.00	2,441,431	2,020,133

Appendix Figure C-7

Number of Records making up the Mutually Exclusive Populations Selected from the IDSHTDSK data sets

ID num	Selection criteria: MATH_C	READ_C	MATH_T	READ_T	num of records	percent of total
1	>0	=0	>0	=0	1,262	0.87
2	>0	=0	=0	>0	277	0.19
3	=0	>0	>0	=0	200	0.14
4	=0	>0	=0	>0	2,833	1.96
5	>0	>0	>0	=0	1,644	1.14
6	>0	>0	=0	>0	4,349	3.02
7	>0	=0	>0	>0	4,126	2.86
8	=0	>0	>0	>0	9,631	6.68
9	>0	>0	>0	>0	59,702	41.42
10	=0	=0	>0	=0	1,749	1.21
11	=0	=0	=0	>0	4,821	3.34
12	=0	=0	>0	>0	53,554	37.15
			Totals		144,148	99.98

ID num	m a t h skills	tests	r e a d i n g skills	tests
1	8,418	1,499	0	0
2	1,540	0	0	326
3	0	202	959	0
4	0	0	11,969	3,243
5	14,831	1,682	10,062	0
6	29,831	0	25,847	5,954
7	23,004	4,336	0	6,067
8	0	10,094	41,720	14,659
9	555,963	64,526	366,690	93,526
10	0	1,810	0	0
11	0	0	0	6,316
12	0	55,531	0	81,240
Totals	633,587	139,680	457,247	211,331

Appendix Figure C-8

TSM11 7:25 thu 6 jun

IDSHSK
(record exists if and only if number of math skills > 0 or number of reading skills > 0 or both)

STUDENT NUMBER	GRADE LEVEL	NUMBER OF MATH SKILLS	NUMBER OF READING SKILLS

Appendix Figure C-9

IDSHTD
(record exists if and only if number of math tests > 0 or number of reading test > 0 or both)

STUDENT NUMBER	GRADE LEVEL	# of math tests			# of reading tests		
		at/above	below	total	at/above	below	total

Appendix Figure C-10

by STUDENT
Each record exists if and only if:
[total # of math tests >0] OR [total # of reading tests >0]

GRADE LEVEL	# of math tests			# of reading tests			# of math skills		# of reading skills	
	at/above	below	total	at/above	below	total	current	curr+prior	cumulative	current curr + prior	cumulative

Appendix Figure C-11